University of Dundee

DOCTOR OF PHILOSOPHY

Post-Replicative Resolution of Under-Replication

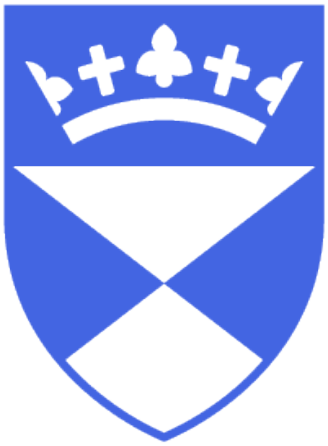
Carrington, James T.

Award date:
2017

Post-Replicative Resolution of Under-Replication

Jamie T. Carrington

Ph.D Molecular and Cellular Biology



School of Life Sciences
The Centre for Gene regulation & Expression
University of Dundee
UK

Supervisor: Professor J. Julian Blow

September 2017

Table of contents

Acknowledgements

I would like to thank Professor Julian Blow for his supervision and guidance throughout my PhD and for providing me with such an interesting project. Additionally, I would also like to thank Dr Alberto Moreno for his day-to-day guidance when I joined the lab. I also thank my other two rotation supervisors, Professor Nicola Stanley-Wall and Dr Kim Dale for two great projects.

I also thank my thesis examiners Professor Jiri Lukas and Dr Constance Alabert, and my convenor Professor Anton Gartner.

My thanks also go out to the rest of the lab. Especially my office mates, Dr Peter Gillespie and Dr Kevin Creavin, for extremely interesting and useful discussions and for all the fun outside of the lab. Additionally, I would like to thank all members of the GRE for all their helpful discussions, as well as to the friends I have made here.

I thank the microscopy and flow cytometry services, which I used extensively, for the highly technical training and very helpful discussions for troubleshooting and experimental design. Which includes Dr Graeme Ball, Dr Paul Appleton, Dr Sam Swift and Mr Calum Thomson, as well as Dr Rosemary Clarke and Ms Arlene Whigham. I would also like to thank the staff who work in the kitchen services, the Garland café and the cleaners.

Special thanks go to Bella Maudlin, who has been of great support for me during my PhD, as well as my parents and Bella's parents and the dogs.

Finally, I thank the Wellcome Trust for funding my PhD, and the University of Dundee for hosting me.

Declaration

I declare that I am the author of this thesis and that all the work presented herein is my own work unless otherwise stated in the main text and/or figure legends. All literature consulted in the writing of this thesis is referenced in the bibliography. This thesis has not been previously submitted or accepted for a higher degree.

Jamie T Carrington

Abstract

The evolutionary pressure to prevent re-replication by inactivating licensing during S phase leaves higher-eukaryotes with large genomes, such as human cells, vulnerable to replication stresses. Origins licensed in G1 must be sufficient to complete replication as new origins cannot be licensed in response to irreversible replication fork stalling. Interdisciplinary approaches between cellular biology and biophysics predict that replication of the genome is routinely incomplete in G2, even in the absence of external stressors. The frequency of converging replication forks that never terminate due to irreversible stalling (double fork stall), which result in a segment of unreplicated DNA, was modelled using high quality origin-mapping data in HeLa and IMR-90 cells. From this, hypotheses were generated that related an increase in unreplicated segments of DNA to reduced functional origin number. Presented in this thesis is the confirmation of this relation by quantifying chromosome mis-segregation and DNA damage responses when origin number was reduced using RNAi against licensing factors. The number of ultrafine anaphase bridges and 53BP1 nuclear bodies are in remarkable concordance with the theoretical predictions for the number of double fork stalls, indicating that cells are able to tolerate under-replication through such post-replicative cellular responses. 53BP1 preferentially binds to chromatin associated with large replicons, and functions synergistically with dormant origins to protect the stability of the genome. Additional candidates, inspired by common fragile site research, have also been characterised as responders to spontaneous under-replication, and include FANCD2 and MiDAS, which function in early mitosis to facilitate completion of replication before cells enter anaphase. In conclusion, a series of mechanisms that sequentially function throughout the cell cycle protects the stability of the human genome against inevitable spontaneous under-replication brought about by its large size.

List of abbreviations

9-1-1	RAD9, HUS1, RAD1
ALT	alternative lengthening of telomeres
ARS	autonomous replication sequence
BIR	break-induced repair
bp	base pair
BrdU	bromodeoxyuridine
CFS	common fragile site
CKI	CDK inhibitor
CMG	CDC45, MCM2-7, GINS
DDR	DNA damage response
DFS	double fork stall
DSB	double-strand break
dNTP	deoxynucleotide triphosphate
EdU	5-ethynyl-2′-deoxyuridine
GINS	*Go-Ichi-Ni-San* (Sld5, Psf1, Psf2, Psf3)
HR	homologous recombination
HU	hydroxyurea
MiDAS	mitotic DNA synthesis
MRN	MRE11, RAD50, NBS1
NHEJ	non-homologous end-joining
OCCM	ORC, CDC6, CDT1, MCM2-7
ORC	origin recognition complex
PI	propidium iodide
pre-IC	pre-initiation complex
pre-RC	pre-replicative complex
RNAi	RNA interference
RPC	replisome progression complex
SCF	SKP, Cullin, F-box
siRNA	short interfering RNA
SNS	short nascent strand
TLS	translesion synthesis
UFB	ultrafine anaphase bridge

Chapter 1

Introduction

1.1) The Cell Cycle

Cells function to proliferate the genetic information stored within their DNA by first replicating their genome and second dividing into daughter cells, each with at least one copy of the genome. These two processes are conserved across all life and are described as the cell division cycle, which in eukaryotes is composed of four sequential phases. The genome is replicated in the synthesis phase (S phase) and the cell divides in mitosis (M phase), which are separated by two growth phases: G1 follows mitosis and G2 follows S phase. The cell cycle is tightly regulated and varies between different organisms. In this thesis, the focus is on the human cell cycle, but it is highly conserved amongst other mammals and even well conserved in yeasts.

The human cell cycle is regulated by the oscillations of activity by a family of serine/threonine kinases known as CDKs (cyclin dependent kinases). There are 20 known mammalian CDKs (CDK1 through CDK20), however, a large number regulate transcription or are tissue specific and are not main players in the progression of the cell cycle (Malumbres, 2014, Malumbres and Barbacid, 2005). The regular mammalian cell cycle is regulated by CDK1, CDK2, CDK4 and CDK6. These kinases promote transitions through the cell cycle phases but, by definition, must be associated with a regulatory subunit for correctly timed activity. The regulatory subunits are known as cyclins and are themselves regulated by sequential proteolysis, which is the principal source of the oscillatory activity of CDKs. Like CDKs, there are a number cyclins with specialised roles, but the core of a normal mammalian cell cycle is regulated by the protein levels of CyclinA, CyclinB, CyclinD and CyclinE. The activity of different combinations of CDK•Cyclin heterodimers defines the phases of the cell cycle, and is summarised below.

Cells are born into G1, which follows from mitosis. CDK1 activity promotes mitosis and drives progression through it, so must be inhibited to reverse physiological changes that are required for mitosis, such as chromatin condensation, nuclear envelope breakdown and spindle assembly. CDK1 activity in mitosis requires CyclinB, and so repression of mitotic CDK1 activity is brought about by

proteolysis of CyclinB. CyclinB contains a domain called the KEN box, which is recognised by CDH1 (CDC20 homologue 1) and activates the APC/C (anaphase promoting complex/cyclosome). The APC/C is an E3 ubiquitin ligase that polyubiquitylates targets, marking them for destruction by the proteasome. CDK4 and CDK6 promote the G1/S transition, and are also inhibited in early G1 by CDK inhibitors (CKIs) the INK4 family of proteins: $p16^{INK4A}$, $p15^{INK4B}$, $p18^{INK4C}$ and $p19^{INK4D}$. APC/C^{CDH1}- and INK4-mediated inhibition ensure that there is little to no CDK activity in early G1 cells. The cell cycle phase transitions to come (G1/S and G2/M) are dependent upon CDK activity, so the drop in CDK activity that ends mitosis and is then resumed to commit cells into the next cell cycle is the primary oscillation of CDK activity. The cell monitors its environment before committing to the next cell cycle. CyclinD promotes the G1/S transition and is itself promoted by growth factors/mitogens. CyclinD complexes with CDK4 and CDK6, which when active, phosphorylate pRb (retinoblastoma protein). pRb, by its binding, inhibits the transcription factor family E2F which promote the expression of many genes, including CyclinE and proteins required for DNA replication. CDK-dependent phosphorylation of pRb reduces its ability to repress E2F-driven transcription. CyclinE binds and activates CDK2 which further phosphorylates pRb. Hyperphosphorylated pRb unbinds E2F, further increasing expression of S phase cyclins and creates a positive feedback loop which ensures the transition to S phase is fast and irreversible. Another family of CKIs further contributes to the fast and irreversible transition between G1 phase CDK activity and S phase CDK activity, called CIP/KIP (CDK interacting protein/kinase inhibitory protein) and includes as its members $p21^{CIP1/WAF1}$, $p27^{KIP1}$ and $p57^{KIP2}$. These three proteins promote the stable interaction between CDK4/6 and CyclinD and inhibit the association of CDK2 with CyclinA/E in G1. CIP/KIP proteins are sequestered to CyclinD, reducing their ability to inhibit CDK2•CyclinA/E after commitment to the cell cycle. High levels of CDK2•CyclinE promoted by E2F overwhelms and antagonises $p27^{KIP1}$ by phosphorylating it, marking it for destruction by a second E3 ubiquitin ligase, the SCF complex (SKP, Cullin, F-box). This second layer of regulation further ensures that G1-CDK activity is mutually exclusive with S-CDK activity, and is irreversible. A third layer of control is the CDK phosphorylation of CDH1. APC/C^{CDH1} stabilises G1 by marking

Cyclins for proteolysis, but as CDK activity accumulates at the G1/S transition, CDH1 becomes phosphorylated which marks it for eventual proteolysis via the SCF complex, and the APC/C can no longer ubiquitylate Cyclins. Once cells have transitioned into S phase and $p27^{KIP1}$ is proteolysed, CyclinA can bind to CDK2, the activity of which increases through S phase. CDK2•CyclinE activity promotes the transition from G1 phase to S phase, but CDK2•CyclinA takes over once S phase begins. CyclinE is marked for proteolysis by SCF to prevent re-replication (more on this later). As cells move from S phase to G2 phase, CyclinA associates with CDK1 and contributes to promoting the G2/M transitions. However, the main source of mitotic CDK activity is from CDK1•CyclinB. CyclinB expression starts in S phase and builds until the end of mitosis. However, CDK1•CyclinB remains inactive due to phosphorylation by the CDK inhibitor WEE1. The antagonistic phosphatase for WEE1 is CDC25, the activity of which is repressed throughout S and G2 phases by SCF-mediated proteolysis. CDK1 phosphorylates WEE1 and CDC25, but the effects are opposed: WEE1 is inhibited and CDC25 is activated. WEE1 is highly sensitive to inhibitory phosphorylation, so when CDK1 activity has reached a sufficient level, WEE1-mediated inhibition of CDK1•CyclinB stops and CDC25-mediated activation begins. Additionally, the specificity of SCF changes from CDC25 to WEE1. Together, these interactions form a bistable switch which ensures fast and irreversible activation of CDK1•CyclinB. Once cells enter mitosis, CDK1 phosphorylates many targets that bring about chromosome condensation, breakdown of the nuclear envelope and spindle formation. CDK1•CyclinB also phosphorylates the APC/C, thereby activating it, and phosphorylates CDH1, thereby deactivating it. When microtubules attach to the kinetochores, the MAD1/2 heterodimer is displaced and CDC20 is activated. One unattached kinetochore is sufficient to repress CDC20, ensuring anaphase does not initiate until ready, reducing the possibility of mis-segregation and aneuploidy. Like CDH1, CDC20 recognises D-box motifs, and coordinates APC/C-mediated ubiquitylation. APC/C^{CDC20} targets CyclinB and securin for proteolysis. Securin binds and inhibits separase, which cleaves cohesin, meaning that APC/C^{CDC20} activity allows the separation of sister-chromatids during anaphase. As APC/C^{CDC20} inhibits its own

promoter, its activity drops after sufficient CyclinB proteolysis, and CDK1•CyclinB repression results in more active CDH1. CDC20 contains a KEN box motif and so APC/C^{CDH1} targets CDC20 for destruction, thus displacing APC/C^{CDC20} activity. As APC/C^{CDH1} does not target G1 and S phase cyclins for proteolysis, the switch from APC/C^{CDC20} to APC/C^{CDH1} ensures that cells irreversibly exit mitosis and enter G1 (Malumbres and Barbacid, 2005, Malumbres, 2014, Sherr and Roberts, 1999, Murray, 2004, Sanchez and Dynlacht, 2005).

The above summary is an idealised version of the mammalian cell cycle, however, there is a lot of redundancy built-in, where most CDK/Cyclin complexes can function redundantly and can be substituted for each other. For example, human and *Xenopus* CDK2 was discovered by cloning an uncharacterised kinase or a cDNA screen, respectively, and were able to complement a *S. cerevisiae* CDK1 mutant (*cdc28*) (Elledge and Spottswood, 1991, Paris *et al.*, 1991). Mouse models which are null for CDK4, CDK6 or CDK2 are viable, as are CDK2/CDK6 double null mutants, and CDK2 null mice have no detectable phenotype in their somatic cells (summarised in Malumbres & Barbacid, 2005). In fact, a mouse model lacking all interphase CDKs develop to mid-gestation, showing that CDK1 is sufficient to drive the mammalian cell cycle (Santamaria *et al.*, 2007). Embryonic fibroblasts derived from this mouse model proliferate, are immortal, and CDK1 binds all the cyclins required for the cell cycle (Santamaria *et al.*, 2007). Additionally, CyclinD and CyclinE are dispensable for mouse development (Geng *et al.*, 2003, Kozar *et al.*, 2004). Given the redundancy and number of CDK/Cyclin complexes, it was argued that all CDK heterodimers are essentially synonymous, but are variably regulated so that they can coordinate complex processes by when and where they are active, as opposed to having biochemically distinct properties for each heterodimer that define a specific set of substrates. To test this, the nuclear localisation signal of CyclinE was recombined with CyclinB, and the recombinant protein was added to *Xenopus* egg extracts depleted of Cyclins. Wild-type CyclinB induced mitosis, but the recombinant protein was able to induce DNA replication, showing that the fundamental difference between CyclinE and

CyclinB in interphase cells is that CyclinE localises to the nucleus and CyclinB does not (Moore *et al.*, 2003).

1.2) DNA replication

The purpose of eukaryotic DNA replication is to duplicate one copy of the genome such that they can be segregated into two separate copies that are nearly identical to the original. Eukaryotic genomes are frequently large, and distributed over multiple chromosomes. Coordination of DNA replication is a significant challenge, and is thus a highly regulated cellular process. Initiation of classical DNA replication defines the beginning of S phase, and is mostly complete when S phase transitions into G2. DNA replication is regulated in space as well as time. During mitosis, the nuclear envelope is broken down so that the duplicated genome can be segregated. APC/C^{CDH1} facilitates exit from mitosis by marking CyclinB for proteolysis, which restores the interphase characteristics of nuclei, including re-establishment of the nuclear envelope. As touched upon previously, the cell uses the boundary between the nucleoplasm and the cytoplasm as a layer of regulation; translation of mRNA into proteins only takes place in the cytoplasm, with a nuclear localisation sequence (NLS) required for a protein to enter the nucleus.

1.2.1) Summary of DNA replication

DNA is double-stranded, and one strand complements the other. In 'the most beautiful experiment', Meselson and Stahl demonstrated that DNA replication is semi-conservative by growing *E. coli* in media that contained heavy ^{15}N and periodically took samples, extracted the DNA and fractionated them in a caesium chloride density gradient. After one cell cycle, the DNA settled at a density halfway in between 'light' ^{14}N DNA and 'heavy' ^{15}N DNA, forming a light/heavy fraction. With each successive cell cycle, the light/heavy fraction transitioned into a heavy/heavy fraction (Meselson and Stahl, 1958). Semi-conservative replication means that a double-stranded molecule of DNA (dsDNA) must be separated before it is replicated, into two single-stranded molecules (ssDNA), each of which can then be used as a template for DNA synthesis. Unwinding of dsDNA into two complementary ssDNA molecules is carried out by a class of proteins called helicases. Eukaryotes

have a number of helicases, many with specialised roles, but canonical DNA synthesis is facilitated by the heterohexamer MCM2-7 (minichromosome maintenance 2 through 7), the catalytic core of the replicative helicase (Tye, 1999, Labib and Diffley, 2001, Bell and Dutta, 2002). Forty MCM genes were discovered in a genetic screen for maintenance of an artificial minichromosome with an origin in *S. cerevisiae* (Maine *et al.*, 1984). The six MCMs that the helicase is composed of are paralogues that are highly conserved across all eukaryotes (Chong *et al.*, 1996, Forsburg, 2004), and form a toroid shape when complexed (Bochman and Schwacha, 2009). MCM2-7-mediated DNA unwinding is processive and moves along the DNA in one net direction, making the nascent ssDNA available as a template for DNA synthesis. DNA is synthesised *in situ* one base at a time on the template by DNA polymerases, complexes of proteins that incorporate and covalently bond dNTPs (deoxyribonucleotide triphosphates) to the 3' end of an existing complement to the template. This means that the processivity of DNA polymerases is strictly 5' to 3', and must start from the interface between a nucleic duplex and ssDNA. As MCM2-7 has separated dsDNA into two ssDNA molecules, there is no duplex that a polymerase can initiate DNA synthesis from, so a specialised polymerase called primase synthesises a short sequence of RNA that complements a section of the ssDNA template, from which DNA polymerase can start synthesising. The two strands of DNA in a dsDNA molecule are antiparallel, so one strand must be synthesised in the 5' to 3' direction, and the other must be 3' to 5'. The strand of DNA synthesised in the 3' to 5' direction need only be primed once and DNA polymerase can synthesis the complement DNA in the 5' to 3' direction, the same direction as the helicase, and is known as the leading strand. However, the other strand cannot be replicated in the same orientation, and so is replicated in short sections as the nascent ssDNA is made accessible by MCM2-7 helicase activity, known as the lagging strand. These short fragments of DNA are called Okazaki fragments and are 100-200 bp in length, with each one requiring priming by primase. DNA polymerisation in the lagging strand starts from the primer that complements nascent ssDNA, in the opposite direction of helicase, and stops at the previously synthesised primer. The lagging strand polymerase displaces the 5' end of the previous Okazaki fragment creating a single-stranded 5' flap, which is then removed by FEN1 (flap endonuclease 1).

The single-stranded nick between Okazaki fragments is then ligated together by DNA ligase 1 (CDC9), completing synthesis of the lagging strand. The eukaryotic replisome contains three different polymerases, Pol α, Pol δ and Pol ε. It was widely accepted that leading strand synthesis was exclusively carried out by Pol ε and the lagging strand by Pol δ, however, recent evidence in yeasts indicates that Pol δ can support leading strand synthesis, albeit at a slower rate, and is replaced by pol ε if it is available (Daigaku *et al.*, 2015, Yeeles *et al.*, 2017, Georgescu *et al.*, 2014). Pol α forms a complex with PRIM1 and PRIM2 subunits of primase and synthesises DNA immediately after the primer, and is quickly taken over by pol δ/ε. (Forsburg, 2004, Bell and Dutta, 2002, Bochman and Schwacha, 2009). PCNA (proliferating cell nuclear antigen) is a homotrimeric sliding clamp that encircles nascent DNA and is loaded by the clamp loader RFC (replication factor C). PCNA topologically tethers pol δ and pol ε to their substrates and enhances their processivity, and is described as a 'docking bay' given the number of factors it binds and stabilises at the fork. For example, FEN1 and ligase 1 bind PCNA which enhances their activity at the lagging strand. When pol δ completes the synthesis of an Okazaki fragment, it is recycled and docked upstream to a newly loaded PCNA (Pomerantz and O'Donnell, 2007).

The above model describes some core components of a single replication fork replicating an arbitrary strand of dsDNA. However, the structure of DNA and of the genome complicates matters. An origin of replication is a genomic locus from where two replication forks are initiated, which replicate DNA in opposite directions upstream and downstream of the origin. As the forks emerge from an origin they expand a replication bubble, as the 5' end of the leading strand of both forks corresponds to the 3' end of the lagging strand of the other fork. The human genome is distributed over 46 linear chromosomes, with origins of replication distributed throughout (the distribution and regulation of origin placement will be introduced later). Bidirectional extension of replication forks from adjacent origins necessitates the activity of topoisomerases, as DNA is a double helix and unwinding at both ends applies torsional stress to the DNA.

1.2.2) Replication of chromatin

DNA is organised into a higher-order structure, called chromatin. ~150 bp of DNA is wrapped twice around a histone octamer, termed a nucleosome, with 50-100 bp of 'linker' DNA between nucleosomes. To allow timely progression of a replication fork, nucleosomes must be dynamically disassembled downstream of the fork and reassembled upstream of it on the newly synthesised DNA. This was neatly demonstrated with an *in vitro* system, where the rate of replication of nucleosomal DNA by a reconstituted eukaryotic replisome was significantly increased by addition of purified FACT (facilitate chromatin transcription), a histone chaperone (Kurat *et al.*, 2017). Nucleosome disruption is highly localised to the replication fork, and is thought to be brought about by either direct collision with the replicative helicase, or positive supercoiling in front of the fork. Additionally, CDK2•CyclinA is recruited to replication forks by interacting with PCNA and MCM7, and phosphorylates histone H1 which increases chromatin decompaction (Alabert and Groth, 2012). Histone chaperones FACT and ASF1 are associated with the replication fork (FACT directly binds MCM2-7 and moves with the fork), and function to recycle H3-H4 subunits of disrupted nucleosomes. Newly synthesised histone subunits H3-H4 are established on the nascent strands by CAF1 (chromatin assembly factor 1), with ASF1, and recruited to the fork by PCNA (Alabert and Groth, 2012, Falbo and Shen, 2006). Histones are extensively post-translationally modified, commonly by methylation, acetylation, phosphorylation or ubiquitylation. These modifications regulate a number of processes, some of which are inherited from one generation to the next, and are thus epigenetic. For example, gene expression is partly regulated by epigenetic histone modifications, if all disrupted nucleosomes on the parental strand are retained in both daughter strands, then the epigenetic marks that regulate gene expression will be diluted amongst the newly synthesised nucleosome components. It is observed by chromatin precipitation that specific histone modifications are restored as an ongoing process through the cell cycle up until G1, and it is suggested that some histone modifiers are recruited to their modification and propagate that same modification to nearby nucleosomes, such as PRC2 at H3K27me3, thus restoring the pre-replicative concentration of histone modifications (Alabert and Groth, 2012). It is also thought that

PCNA rings remain on newly synthesised DNA in the wake of the replication fork and can recruit DNMT1 (DNA methyltransferase 1) to re-establish DNA methylation at CpG sites (Alabert and Groth, 2012).

There are a number of factors that associate with chromatin that have the potential to impede replication fork progression, such as DNA damage repair or transcription, and will be introduced later.

1.2.3) Spatial regulation of DNA replication

In higher eukaryotes, origins of replication are organised into clusters at two main levels. Firstly, a 'replication unit' which can contain several origins and span 50-100 kbp (Fragkos *et al.*, 2015). Replication units contain several origins which have the potential to be activated, but only one of which is stochastically activated, implying that activation of one origin represses the activation of the rest in the replicating unit (Lebofsky *et al.*, 2006). This phenomenon is called 'negative origin interference' and is thought to be mediated by CHK1 and CHK2 signalling (Brewer and Fangman, 1993, Lebofsky *et al.*, 2006, Fragkos *et al.*, 2015), but it was also shown that passive replication of neighbouring origins is sufficient to explain this phenomenon (Blow and Ge, 2009). This affords flexibility; if one part of the replication unit cannot initiate replication, due to localised DNA damage, then another origin within the unit can compensate. The second level of origin organisation is replication clusters, which consists of several replication units, and can span up to 1 Mbp in length. Clusters are thought to correspond to replication foci, or replication factories, which are made visible by incorporation of a thymidine analogues that can be fluorescently labelled, such as BrdU (5-bromo-2'-deoxyuridine) during S phase (Berezney *et al.*, 2000, Jackson and Pombo, 1998). Origins within a cluster are synchronously activated, by 'positive replication interference' (Marheineke and Hyrien, 2004). There are ~800 replication foci activated at the peak of S phase, each containing ~2-5 active activated origins (Jackson and Pombo, 1998). This was more recently confirmed by comparing the number of replication foci in HeLa cells expressing GFP-PCNA

between wide-field microscopy (which applies to ~800 replication foci quantified previously) and super-resolution microscopy, producing a ratio of ~1:5, and calculating the number of active replicons and finding that under the super-resolution microscope one replication foci approximately corresponds to one replicon (Chagin *et al.*, 2016). It is thought that replication units correspond to chromatin loops, and that cohesion brings several chromatin loops together to form a replication cluster (Guillou *et al.*, 2010, Fragkos *et al.*, 2015). Replication foci are stable units of chromosome structure that last several cell cycles (Jackson and Pombo, 1998) and do not move, merge or divide (Leonhardt *et al.*, 2000), indicating that they are specialised sites that facilitate replication, as opposed to random chromatin coalescence.

1.2.4) Temporal regulation of DNA replication

All the origins that are activated in a given S phase are not activated simultaneously. The eukaryotic genome is replicated in segments which are activated in a predictable order. This is evident when imaging replication foci at time-points throughout S phase. Human cells synchronised and released from the G1/S boundary were pulsed with BrdU, which when immunolabelled form distinctive patterns of replication foci that are highly reproducible. At the onset of S phase, replication foci quickly appear throughout the nucleus, by mid S phase the signal is strongest at the periphery of the nucleus and in late S phase a few large bodies of BrdU incorporation can be seen nearer the middle of the nucleus. Five distinct replication foci patterns emerge and occur in the same order with similar timing, and is conserved among higher eukaryotes (O'Keefe *et al.*, 1992). The replication timing programme is likely a consequence of the complex chromosomal structures and folds, co-regulation with transcription and limiting concentrations of replication factors. Early-replicating segments occupy domains within the nucleus that are gene-rich and associated with high levels of transcription, and euchromatin, and are also enriched with replication origins. It is possible that the decompacted state of euchromatin is more permissible to the formation of replication origins, thus reducing the initiation time of replication at those loci. Late replication is associated with repetitive elements, heterochromatin and lamina, at the nuclear periphery and are origin-poor (Rhind and

Gilbert, 2013, Fragkos *et al.*, 2015). The primary DNA sequence can also correlate with early replicating segments. Early replicating segments are LINE-poor (long interspersed nuclear elements, retrotransposons that make up ~20% of the human genome) and isochore-rich (high GC% regions) (Rhind and Gilbert, 2013). Chromatin state and primary DNA sequence correlate with replication timing but are not sufficient to define it. The greatest determinant yet discovered for replication timing is the spatial organisation of chromosomes. Chromosome conformation capture is a sequencing technique that can determine the spatial organisation of interphase chromatin, whereby DNA is crosslinked, digested, self-ligated, reverse-crosslinked and then sequenced. This technique identifies associated domains of chromatin where the DNA interacts internally within the domain, but less so with DNA within other domains. The boundaries between these self-interacting domains bear a striking resemblance to the boundaries between early replicating and late replicating segments, indicating that replication timing is regulated by nuclear compartments defined by chromatin folding (Ryba *et al.*, 2010, Rhind and Gilbert, 2013). Modelling the replication timing programme in the context of euchromatin and heterochromatin convincingly simulates the patterns of replication foci seen throughout S phase. Löb *et al.* 2016, created a model in 3D space whereby origins in euchromatin can stochastically activate at a high probability, and so activate randomly but early in S phase, and then with a 'domino-like affect' activate origins in adjacent self-interacting domains. The last DNA to be replicated is associated with heterochromatin, and visually mimics the S phase replication foci pattern timings originally observed by O'keefe *et al.* 1992 (Lob *et al.*, 2016).

1.2.5) Replication origins and origin 'licensing'

Although the factors that are recruited to and define an origin of replication are highly conserved among eukaryotes, the properties of the genomic loci to which those factors are recruited differ between yeasts and metazoans. The majority of *S. cerevisiae* origins contain a consensus sequence, which is an essential component of the conserved ARS (autonomous replication sequence) which contains three non-essential motifs, all together spanning just over 100 bp. ARSs provide binding

sites for replication factors, and are AT-rich, thus requiring less energy to melt the DNA duplex at the initiation of replication (Bell, 1995). However, in higher eukaryotes, such as in humans, no such replicative elements have been discovered. Certain properties of the primary sequence and chromatin correlate with replication origin positions, but are not sufficient to define them. Genomic loci at replication origins in human cells are often GC-rich, have uncompacted chromatin, as facilitated by histone acetylation, and have a G-quadruplex ~220 bp 3' of the site of initiation. Additionally, replication origins are excluded from sites of active transcription (Gilbert, 2001, Prioleau and MacAlpine, 2016). A minority of human replication origins do have sequence-defined sites, such as the origin within the β-globin locus (Aladjem, 2004).

Whether origins are at defined genomic loci or stochastically distributed within a domain, there is a set series of events that make a genomic locus competent for DNA replication, the product of which is a complex of proteins on DNA called the pre-replicative complex (pre-RC). This process is known as replication origin 'licensing' and is highly conserved among eukaryotes, with only a few differences between *S. cerevisiae* and vertebrates. The first step is the binding of DNA by ORC (origin recognition complex), which is necessary but not sufficient for eventual origin activation. ORC is a hexameric (ORC1 through ORC6) AAA+ ATPase, and ORC must be bound to ATP for the complex to bind to the DNA (Bell and Dutta, 2002). However, reconstitution of licensing with purified *Xenopus* egg extract showed that ORC binding to DNA can take place in the presence of ADP (Gillespie *et al.*, 2001). In humans, the complex contains a stable core of ORC2-5, that weakly interacts with ORC1 and ORC6 (Masai *et al.*, 2010). ORC1 is synthesised from the beginning of G1 and then proteolysed via SCF^{SKP2}-mediated ubiquitylation during S phase (Mendez *et al.*, 2002, Masai *et al.*, 2010). ORC2-5 binds to the DNA-bound ORC1•ATP, forming ORC1-5. ORC6 is not essential for licensing in *Xenopus* or HeLa cells (Vashee *et al.*, 2001, Gillespie *et al.*, 2001), but is required for DNA replication (Li and Herskowitz, 1993). ATP-bound CDC6 binds to DNA-bound ORC, and is another AAA+ ATPase, highly related to ORC1 (Bell and Dutta, 2002). Cryo-electron microscopy (EM) indicates that ORC1-5 and CDC6•ATP form a ring structure of a similar size to MCM2-7 (Speck *et al.*, 2005). The DNA-binding motifs of ORC1-5 are orientated towards the

lumen of the ring, indicating ORC1-5•CDC6 forms a ring around DNA. In *S. cerevisiae*, the ATPase activity of CDC6 destabilises its interaction with ORC, which is suppressed by binding origin DNA (Yardimci and Walter, 2014). In *S. cerevisiae*, DNA-bound ORC•CDC6 provides a platform for the loading of CDT1-bound MCM2-7. Like ORC and CDC6, MCM2-7 is a member of the AAA+ ATPase family of proteins, and the heterohexamer forms a ring shape with a positively charged central channel that is wide enough to contain dsDNA, and was confirmed to do so by EM of reconstituted licensing with *S. cerevisiae* proteins (Remus *et al.*, 2009, Evrin *et al.*, 2009). The interface between MCM2 and MCM5 in the heterohexamer is open, and was demonstrated to be a 'gate' through which DNA passes, as MCM2-7 loading was blocked by inducing the dimerisation of fusion proteins MCM2-FRB and MCM5-FKBP (Samel *et al.*, 2014). The six members of the MCM2-7 complex are parallel with respect to the N- and C-termini of each protein (Yardimci and Walter, 2014). CDT1 (Cdc10-dependant transcript 1) forms a soluble heptamer with MCM2-7 in *S. cerevisiae* (Tanaka and Diffley, 2002), that docks at ORC•CDC6 as a stable unit (Yardimci and Walter, 2014). Cryo-EM of the OCCM structure (ORC•CDC6•CDT1•MCM2-7) on DNA indicates that ORC•CDC6 behaves like a clamp-loader for MCM2-7•CDT1, whereby the helical conformation of ORC•CDC6 causes MCM2-7 to adopt a similar conformation once bound, which opens that MCM2-5 gate (Sun *et al.*, 2013). CDT1 could cause the open conformational change required for MCM2-7 to wrap around DNA, as it binds primarily to MCM2, and partially to MCM5 (Sun *et al.*, 2013). The OCCM forms *in vitro* with a non-hydrolysable form of ATP, but is salt-sensitive (Masai *et al.*, 2010). In the OCCM, CDC6 is closest to the C-terminus of MCM3, which is required for the loading of MCM2-7 to ORC•CDC6, orientating the N-terminus side of MCM2-7 away from the ORC (Frigola *et al.*, 2013). Soon after OCCM formation, CDT1 dissociates dependent on ATP-hydrolysis by MCM (Coster *et al.*, 2014), which forms the OCM. The OCM is a platform to load a second MCM2-7•CDT1 heptamer. The mechanism of this step is not yet clear, but is known to require a second molecule of CDC6, and ATP hydrolysis, and results in head-to-head contact between the two MCM2-7 complexes loaded on DNA (Deegan and Diffley, 2016,

Yardimci and Walter, 2014, Remus *et al.*, 2009). The head-to-head orientation of loaded double hexamers is likely the source of the bidirectionality of replication forks away from the origin upon initiation and elongation. Reconstitution of *Xenopus* and *S. cerevisiae* purified proteins demonstrated that, *in vitro*, MCM2-7, ORC1-6, CDT1 and CDC6 were necessary and sufficient to support formation of the pre-RC and to license origins for replication (Gillespie *et al.*, 2001, Remus *et al.*, 2009).

1.2.6) Origin activation

Given the salt-resistance of the MCM2-7 double hexamer on DNA, ORC, CDT1 and CDC6 could be dissociated from DNA, leaving the MCM2-7 double hexamer as the only component of the pre-RC required for origin activation, in a reconstituted *in vitro* system (On *et al.*, 2014, Rowles *et al.*, 1999, Yeeles *et al.*, 2015). Activation of the double hexamer requires phosphorylation events from CDK2•CyclinA/E and from DDK (Dbf4-dependent kinase, DBF4•CDC7). As with CDKs, CDC7 is expressed throughout the cell cycle, but depends on complexing with DBF4 to become an active kinase, and as with CyclinA/E, DBF4 is upregulated during S phase, and therefore CDK and DDK function in parallel to activate licensed origins in S phase. Purified DDK phosphorylates the N-termini of MCM2, MCM4 and MCM6 within the double hexamer, with little to no phosphorylation of their soluble forms (Francis *et al.*, 2009). Phosphorylated MCMs facilitate the recruitment of CDC45 (Jares and Blow, 2000, Sheu and Stillman, 2006) and SLD3/7 (Tanaka *et al.*, 2011). CDK phosphorylation of SLD2 and SLD3 (RECQL4 and Treslin in humans) at the double hexamer, facilitates their interaction with Dpb11 (TOPBP1 in humans, topoisomerase II-binding protein 1) (Zegerman and Diffley, 2007, Tanaka *et al.*, 2007). Phosphorylation of SLD2 also increases its association with the heterotetramer GINS (SLD5, PSF1, PSF2, PSF3). CDC45 and GINS, like MCM2-7, are essential for elongation and colocalise at the replication fork, to form a complex called CMG (CDC45, MCM2-7, GINS) which is the replicative helicase (Moyer *et al.*, 2006, Tanaka *et al.*, 2007). Cryo-EM structures of double CMGs indicate that each MCM2-7 at an origin develops into a CMG prior to MCM separation, perhaps maintaining the CMGs in close proximity

may ensure coupled activation (Costa *et al.*, 2014). Reconstitution of *S. cerevisiae* replication initiation with purified proteins demonstrated that the minimal set of proteins required for the pre-initiation complex (pre-IC) are Mcm2-7, Cdc45, GINS, Sld2, Sld3/7, Dpb11 and pol ε, which were all necessary and sufficient (Yeeles *et al.*, 2015). MCM10 is not required for CMG formation in this *in vitro* system, but is required for recruitment of the ssDNA binding protein RPA (replication protein A), which coats ssDNA to stabilise and protect it from degradation, implying that MCM10 is required for CMG helicase activity (Yeeles *et al.*, 2015). In addition to facilitating the recruitment of pre-IC components, MCM phosphorylation by DDK is thought to convert MCM2-7 from an inactive state in G1 to an active conformation in S phase. For example, the *S. cerevisiae MCM5* mutant *bob1* (P83L) bypasses the requirement of DDK activity for activation of helicase activity, as proline is not a post-translation modification site and is often required for backbone structure in a peptide. *bob1* mimics a conformational change in MCM2-7 that is normally caused by DDK activity (Hardy *et al.*, 1997).

1.2.7) Elongation

Transitioning from the pre-IC to the RPC (replisome progression complex) requires melting of the origin DNA and switching MCM2-7s substrate from dsDNA to ssDNA, clamp loading of PCNA by RFC, recruitment of pol α, checkpoint mediator MRC1, fork stabilisers TOF1 and CSM3, a homotrimer of CTF4, MCM10, topoisomerase I and FACT.

The mechanism of DNA melting in eukaryotic origins is not known. A recent cryo-EM structure of a *S. cerevisiae* MCM2-7 double hexamer purified from G1 chromatin, with a resolution of 3.8 Å, reveals that the central channels of each hexamer are offset and twisted relative to one another, and are likely to cause a kink in the DNA that initiates melting of the duplex (Li *et al.*, 2015). Whether MCM2-7 translocates dsDNA or ssDNA to facilitate unwinding was unknown, with several models proposed (Takahashi *et al.*, 2005), but an experiment in *Xenopus* egg extract demonstrated that strand-specific fork obstruction stopped the progression of MCM2-7 on the translocation strand

(corresponds to the leading strand) but not on the excluded strand (corresponds to the lagging strand), favouring the 'steric exclusion' model over the 'ploughshare' model, and defining MCM2-7 as a 3'-5' ssDNA translocase (Fu *et al.*, 2011). How a salt-resistant MCM2-7 hexamer wrapped around dsDNA develops into a 3'-5' ssDNA translocase is not clear, but the authors speculate that following melting of dsDNA the MCM2/5 gate reopens during S phase and opposite ssDNA molecules are extruded from each hexamer, perhaps aided by MCM10, and the binding of CDC45 and GINS at the MCM2/5 gate (as seen in cryo-EM structures (Costa *et al.*, 2014)) closes the gate and activates the helicase (Fu *et al.*, 2011).

A number of components of the RPC were identified by mass spectrometry of proteins that were pulled out of *S. cerevisiae* cell lysate with purified GINS. Other than components of the CMG, TOF1, CSM3, MRC1, CTF4 and FACT were precipitated, with low levels of topoisomerase I and MCM10 (Gambus *et al.*, 2006). CTF4 forms a homotrimer, and each molecule contains a domain that binds a motif found in multiple proteins called CIP-box (CTF4 interacting peptide), one valency of the trimer binds GINS and the other two can recruit multiple factors to the fork, crucially pol α (Villa *et al.*, 2016, Gambus *et al.*, 2009). CTF4 is also important for establishing cohesion between sister-chromatids (Hanna *et al.*, 2001).

Once the two RPCs of an origin are established, leading and lagging strand synthesis can take place, as described previously, and the forks progress away from the origin bidirectionally, expanding the replication bubble. MCM2-7s on chromatin, that remain inactivated, are displaced from DNA by a replication fork from a nearby origin, a process known as passive replication (Yekezare *et al.*, 2013, Santocanale and Diffley, 1996).

1.2.8) Termination

The complicated DNA structure at converged replication forks must be resolved into two separate duplexes of DNA. A precise, and elegant, mechanism for eukaryotic termination has only recently been proposed.

Firstly, experiments in *S. cerevisiae* and *Xenopus* egg extract revealed that polyubiquitylated MCM7 is enriched at the end of S phase and results in rapid disassembly of CMG, which is dependent on the activity of CDC48/p97 segregase (Moreno *et al.*, 2014, Maric *et al.*, 2014). In yeasts, ubiquitylation of Mcm7 was found to be SCF^{Dia2}-dependant, which was only visible *in vivo* after AID-degron-mediated depletion of Cdc48, indicating that MCM7 proteolysis is rapid following ubiquitylation (Maric *et al.*, 2014). *dia2Δ* yeast cells, and blocked ubiquitylation of MCM7 in *Xenopus* egg extract could not dissemble their CMGs, which accumulated at the end of S phase (Moreno *et al.*, 2014, Maric *et al.*, 2014). It was previously shown that SCF^{Dia2} is tethered to the RPC via a domain that facilitates binding to CTF4 and MRC1 (Morohashi *et al.*, 2009), so it is not clear how SCF^{Dia2} is regulated to target only terminated CMGs for degradation.

Secondly, experiments in *Xenopus* egg extract on plasmid constructs with site-specific replication fork barriers revealed that converging CMGs pass each other without slowing, as opposed to colliding, and continue until the strand along which they translocate interfaces with the excluded strand of the other fork (Dewar *et al.*, 2015). The CMGs do not dissociate from the DNA until the gaps between the leading strand of one fork is ligated to the last Okazaki fragment of the other fork, implying that MCM2-7 is unloaded from dsDNA (Dewar *et al.*, 2015). This is consistent with *in vitro* experiments where purified MCM2-7 was capable of translocating along a 3' ssDNA overhang and continued when MCM2-7 encountered dsDNA (Kang *et al.*, 2012). Dewar *et al.* 2015 speculate that MCM2-7 on dsDNA is the substrate for SCF^{Dia2}-mediated ubiquitylation, differentiating terminated CMGs from active CMGs.

1.2.9) Dormant origins.

A series of observations demonstrate that origin licensing does not always conform to the model described above, and are collectively termed the 'MCM paradox' (Hyrien *et al.*, 2003). The number of MCM2-7 double hexamers loaded onto chromatin in *S. cerevisiae*, *Xenopus* egg extract and human cells is well in excess of the number of initiation events at origins (Blow and Dutta, 2005, Burkhart *et al.*, 1995, Lei *et al.*, 1996, Mahbubani *et al.*, 1997). MCM2-7 double hexamers were also observed to be distributed at significant distances away from the ORC. *In vitro* experiments in *Xenopus* egg extract revealed that the quantity of MCM on synthetic DNA molecules with one ORC increased linearly with the length of DNA, up until ~3 kbp (~35 double hexamers) (Edwards *et al.*, 2002). Immunoprecipitation of crosslinked nucleoproteins with MCM antibodies, followed by shearing and fractionation, revealed that MCM can be on DNA up to ~1 kbp away from an ORC in HeLa cells (~10 double hexamers) (Ritzi *et al.*, 1998). Additionally, sites of DNA synthesis only contain a fraction of the chromatin-bound MCM complexes (Blow and Dutta, 2005, Dimitrova *et al.*, 1999, Krude *et al.*, 1996, Madine *et al.*, 1995). *Xenopus* egg extract can support a normal replication rate after reducing the availability of MCMs to ~1-2 per active origin (Blow and Dutta, 2005, Edwards *et al.*, 2002, Mahbubani *et al.*, 1997, Woodward *et al.*, 2006), and the affinity of CDC6 for origins is reduced once they are licensed (Oehlmann *et al.*, 2004). Together, these observations indicate that a significant proportion of MCM2-7 is loaded onto chromatin but are rarely activated. It is estimated that ~30% of licensed origins are stochastically activated in a given S phase (Blow and Ge, 2009, Blow *et al.*, 2011, Ge *et al.*, 2007, Ibarra *et al.*, 2008, Woodward *et al.*, 2006).

It was hypothesised that the excess origins could be activated in response to insufficient replication from the ~30% of origins that are normally activated (Hyrien *et al.*, 2003, Edwards *et al.*, 2002, Blow and Dutta, 2005). Supplementing *Xenopus* egg extract with increasing concentrations of the CDT1 inhibitor geminin can be used to limit licensing up until replication dynamics are affected, a state termed 'minimal licensing' (Oehlmann *et al.*, 2004, Woodward *et al.*, 2006). Minimally

licensed extracts exhibited a normal frequency of origin activations but are unable to complete replication when treated with low doses of replication inhibitors such as aphidicolin and the ATR (ataxia telangiectasia and rad3 related protein) inhibitor caffeine, where normally licensed extracts could (Woodward *et al.*, 2006). In human cells, a ~50% depletion of licensed origins had no significant effect on the frequency of origin activations compared to control cells. When treated with HU, depleted cells did not activate dormant origins, had a reduced nucleotide incorporation rate and were hypersensitive compared to control cells (Ge *et al.*, 2007).

Taken together, CDC6 has the potential to load multiple copies of MCM2-7 at one ORC. Within an origin cluster, each origin has the potential to be stochastically activated, and when one is it supresses the activation of neighbouring origins by passive replication. When faced with replication stress, and fork progression is slowed or stopped, then other origins within the cluster stochastically activate (Alver *et al.*, 2014, Blow and Ge, 2009). Disruption of replication forks causes a decoupling of the CMG and the polymerases, which increases the exposure of RPA-coated ssDNA that is threaded between them, and is the substrate for ATR recruitment and activation (Branzei and Foiani, 2009). ATR activates CHK1 DNA damage checkpoint kinase, which inhibits the activation of new replication factories but permits the activation of new origins within already activated factories (Alver *et al.*, 2014, Ge and Blow, 2010). As discussed in Blow and Ge, 2009, this model makes evolutionary sense as licensing inefficient dormant origins, that will only be activated where needed and after the majority of replication has already taken place, is less demanding on cellular resources than the addition of efficient origins at the relevant sites (Blow and Ge, 2009). This is supported by experiments in human cells where inhibition of ATR caused unrestrained origin activation which caused depletion of the cellular pool of RPA, and thus the unprotected ssDNA at replication forks developed into DNA damage (Toledo *et al.*, 2013). RPA is rate-limiting on the number of replication forks that can be active at once, so dormant origins which only become active if a nearby efficient origin stalls, will not significant contribute to the total rate.

1.3) The licensing system

The licensing system is an elegant solution to ensure complete replication of the genome without replicating any section more than once (Blow and Laskey, 1988). This is achieved by strictly separating the biochemistry that drives origin licensing from origin activation. The first step is to load MCM hexamers onto chromatin from late mitosis and throughout G1. The second step begins at the G1/S transition, where the loading of MCMs is strictly halted and origins are activated. These two steps must not overlap as the cell would have nothing to biochemically differentiate an origin that requires MCM2-7 for activation, and an origin that has already been replicated. If licensing were permitted to take place in S phase then replicated origins will be relicensed and re-replicated. However, although the licensing system effectively prevents re-replication, it creates additional challenges to ensure that all of the genome is duplicated. As licensing cannot take place in S phase, the origins that are licensed must be sufficient to support replication of the entire genome. In the event of replication stress, no new replication factors can be recruited to rescue replication. The loading of MCM2-7 at an origin licenses it for activity, and when that origin is activated the MCM complexes elongate away from the origin as part of the replicative helicase CMG, thus when S phase begins the origin loses its license, which cannot be regained until after the following mitosis (Blow and Laskey, 1988).

The licensing system is tightly regulated at multiple levels by the cell cycle. Licensing is limited to late mitosis and early G1 largely by the regulation of ORC, CDC6 and CDT1, as they are required for licensing, but not for replication (Blow and Dutta, 2005). The mechanism by which licensing is limited to late mitosis and G1 varies between lower and higher eukaryotes. In yeasts, CDK activity is high from the G1/S transition through to late mitosis, and phosphorylates MCM, ORC1, CDC6 and CDT1 (Arias and Walter, 2007, Blow and Dutta, 2005). Phosphorylation of CDC6 signals it for proteolysis via ubiquitylation by SCF^{Cdc4}, and causes it to form a stable association with mitotic CDKs, disabling its licensing activity, additionally, CDK inhibits CDC6 transcription by blocking the nuclear import of transcription factor Swi5 (Arias and Walter, 2007). Phosphorylation of

MCM2-7 by CDK results in its export from the nucleus (Liku *et al.*, 2005), along with CDT1, due to their association (Tanaka and Diffley, 2002). CDK phosphorylation of Orc2 and Orc6, inhibits ORC function, and the yeast S phase Cyclin Clb5 binds directly to Orc6 after origin activation, inhibiting the formation of a pre-RC (Arias and Walter, 2007, Wilmes *et al.*, 2004). Inhibition of any one of these six processes is not sufficient to cause significant re-replication, nor is inhibition of different combinations of two or three, only when most, or all but one, are inhibited does re-replication take place (Nguyen *et al.*, 2001). The unit of re-replication is the number of origins that are relicensed and activated. In *S. cerevisiae*, there are ~500 origins, so if relicensing is supressed with a 99% efficiency then re-replication will occur from ~5 origins. The redundancy of inhibition of pre-RC components during yeast S phases not only makes repression of re-replication robust, but also combine together to repress even a single origin of re-replication from occurring. For example, two mechanisms that each repress re-replication with a 99% efficiency would combine together to suppress re-replication with a 99.99% efficiency.

In metazoans, licensing is largely restricted to late mitosis and G1 by the inhibition of CDT1 activity. This is achieved by the degradation of CDT1 and inactivation by the coiled-coil protein geminin. The synthesis of CDT1 remains relatively constant throughout the cell cycle but protein levels decline during S phase (Nishitani *et al.*, 2001). Proteolysis of CDT1 is mediated by two different E3 ubiquitin ligases. CDT1 binds to chromatin-bound PCNA via its PIP (PCNA-interacting protein) box, restricting this process to S phase, and is then bound by CUL4-DDB1^{CDT2} which ubiquitylates CDT1, targeting it for destruction (Arias and Walter, 2007, Blow and Dutta, 2005). CDT1 is also phosphorylated by CDK activity in S and G2 phase, which targets it for ubiquitylation by SCF^{SKP2}, and thus proteolysis (Takeda *et al.*, 2005). Geminin binds CDT1 to physically block its interactions with MCM2-7 (Lee *et al.*, 2004, Xouri *et al.*, 2007). Geminin is marked for proteolysis, or inactivation, by the APC/C, which means that geminin activity is reduced in late mitosis and early G1, thus CDT1 is able to function as a licensing factor following mitosis and up until the transition to S phase (Arias and Walter, 2007, Blow and Dutta, 2005). Proteolysis

of CDT1 by SCF^{SKP2} is blocked when geminin binds CDT1, and CUL4-$DDB1^{CDT2}$ activity is switched off after S phase, so CDT1•geminin complexes accumulate from G2 through to mitosis (Arias and Walter, 2007). The APC/C mediates the destruction of geminin, which means that a free pool of CDT1 is immediately available for licensing as cells exit mitosis (Arias and Walter, 2007).

1.3.1) Over-replication

Unprogrammed re-replication can be highly deleterious in eukaryotic cells by increasing gene dosage, causing DNA damage, and reducing genomic stability. Overexpression of licensing factors is a hallmark and prognostic of many cancers (Blow and Gillespie, 2008), but it is not clear whether that is what drove tumorigenesis or is a consequence of a deregulated proliferation. A study of CDT1 and CDC6 expression levels of lung, colon and head-and-neck cancers found no correlation between the >4-fold overexpression of CDT1 and CDC6 with the proliferation marker Ki-67, leading the authors to hypothesise that in the absence of functional p53, re-replication caused the DNA damage and genomic instability that resulted in tumorigenesis (Karakaidos *et al.*, 2004, Liontos *et al.*, 2007).

A mutant *S. cerevisiae* cell line, where CDK regulation of Orc2, Orc6, Mcm2-7 and Cdc6 is blocked, undergoes re-replication and its chromosomes become fragmented (Green and Li, 2005). A Rad9-dependent and Mrc1-independent checkpoint is activated by re-replication, indicating that DNA damage has been caused, rather than stalled forks and replication stress (Green and Li, 2005). Similar experiments in mammalian cells which are geminin, DDB1 or CDT2 mutants (so CDT1 levels remain high in S phase and beyond) exhibit increased ɣ-H2AX foci (double strand breaks, DSBs) and activate CHK1 and CHK2 which downregulate CDK1 and thereby entry into mitosis (Zhu and Dutta, 2006, Lovejoy *et al.*, 2006, Jin *et al.*, 2006, Zhu *et al.*, 2004, Melixetian *et al.*, 2004). Inhibition of CHK1 in these experiments forced cells to enter mitosis and resulted in mitotic catastrophe and apoptosis.

A similar experiment shows that *Xenopus* egg extract supplemented with purified CDT1 causes re-replication and CHK1 activation (Li and Blow, 2005). One way in which re-replication has been proposed to cause DSBs is a head-to-tail collision between a re-replicative fork and the replicative fork from the same origin (Davidson *et al.*, 2006).

1.3.2) Under-replication

In contrast to over-replication, where licensing overlaps with S phase, under-replication can occur when S phase begins before a sufficient number of origins are licensed. Under-replication directly causes genomic instability as fork structures in the DNA are unstable, and can indirectly contribute to genome instability by physically linking sister-chromatids, a structure that must be resolved for errorless segregation during anaphase. A MCM4 hypomorphic mouse model exhibited increased chromosome breaks and micronuclei when treated with aphidicolin, had a reduced ability to activate dormant origins, and were prone to adenocarcinomas (Shima *et al.*, 2007). A similar study found that mice with reduced MCM2 levels also exhibited increased frequencies of micronuclei and γ-H2AX (Pruitt *et al.*, 2007). Overexpression of CyclinE causes a reduction in licensing in human cells (Ekholm-Reed *et al.*, 2004), genomic instability and tumorigenesis in mice (Loeb *et al.*, 2005, Spruck *et al.*, 1999) and triggers the DNA damage response (DDR) and eventual senescence in a primary human cell line (Bartkova *et al.*, 2006).

1.4) Replication fork stalling

In addition to insufficient origin licensing, under-replication can be caused by slowed or stopped replication fork progression.

1.4.1) Fork stalling

The progression of replication forks can be stalled. However, origins are licensed in excess to levels that reduce the impact of fork stalling. Firstly, as previously mentioned, many more origins are licensed than are typically used in a given S phase, and otherwise dormant origins can be

stochastically activated to rescue a replicon from fork stalls. Secondly, more origins are activated, even in the absence of replication stress, than are strictly necessary to complete replication within the time constraint of S phase. For example, initiation events in human cells are a mean of ~100 kbp apart (Conti *et al.*, 2007, Jackson and Pombo, 1998), so the mean number of bases a fork replicates is ~50 kbp, yet each fork can potentially replicate DNA in the order of megabases (Blow and Ge, 2008, Newman *et al.*, 2013). The excess initiation events reduce the total time for complete replication of the genome, but will also mean that a fork stall from one origin can still terminate with the corresponding fork from the neighbouring origin.

There are multiple causes of replication fork stalling, including DNA adducts, single strand breaks, collision with transcription factors, DNA-bound proteins, ssDNA secondary structures and torsional and topological stress (Mankouri *et al.*, 2013). Additionally, exogenous reagents can be used to stall replication forks. Commonly used examples include, aphidicolin, a pol α inhibitor, and hydroxyurea, an inhibitor of ribonucleotide reductase, which results in the depletion of the pool of dNTPs.

1.4.2) Checkpoint signalling

If a sufficient number of replication forks stall, this will be detected by cell cycle checkpoints that either protect the cell from exacerbating damage by transmitting it through complex cellular processes such as S phase and mitosis, irreversibly hold the cell in a state of senescence or activate apoptosis. The two main checkpoint sensors are ATR and ATM (ataxia telangiectasia mutated), both of which belong to the phosphatidylinositol 3-kinase-related kinase family, which phosphorylates S/TQ motifs. Targets of these kinases include transducers CHK1 and CHK2 which acts upon effectors such as p53, DDK, and CDC25, which can trigger cell cycle arrest prior to S phase, mid-S phase and mitosis, respectively (Sancar *et al.*, 2004, Blackford and Jackson, 2017).

ATR-mediated checkpoint activation is associated with replication stress. The main phenomenon detected by the ATR pathway is long tracks of ssDNA coated by RPA. As mentioned previously, decoupling of the replicative helicase, CMG, from the DNA polymerases within a fork cause the ssDNA between the two to become extruded, exposing a track of RPA-coated ssDNA (Branzei and Foiani, 2009). Tracks of RPA recruit ATRIP (ATR-interaction protein) which in turn recruits ATR to ssDNA (Cortez *et al.*, 2001). RPA is also bound by RFC^{RAD17}, which clamp-loads the PCNA-like heterotrimeric 9-1-1 complex (RAD9, HUS1, RAD1), which activates RPA-recruited ATR. ATR activates RPC-component MRC1, which recruits RAD17 and together facilitates phosphorylation of CHK1 (Blackford and Jackson, 2017). Activated CHK1 stabilises MRC1 at the RPC, completing a positive feedback loop that amplifies ATR signalling at a stalled fork. Once activated by phosphorylation, CHK1 is released from the fork and transduces the stress signal to p53, CDC25 and DDK (among others) (Blackford and Jackson, 2017). ATR and CHK1 is in constant feedback with fork progression, and is required for normal fork progression in unperturbed S phases (Petermann *et al.*, 2006). Depletion of ATR activity is mouse embryonic lethal and stops proliferation of cells or cell death in the absence of induced replication stress or DNA damage (de Klein *et al.*, 2000, Brown and Baltimore, 2000).

ATM-mediated checkpoint activation is associated with DSBs. A proportion of stalled replication forks collapse, forming a DSB, either by endonucleolytic cleavage of replication intermediates or if the fork encounters a ssDNA lesion (Yeeles *et al.*, 2013, Berti and Vindigni, 2016). One of the first responders to a DSB is the MRN complex (MRE11, RAD50, NBS1), which binds the DSB ends, and stabilises them at the break. ATM is recruited to the break by MRN, and upon activation phosphorylates local modified histone H2AX, forming ɣ-H2AX. MDC1 binds ɣ-H2AX, which facilitates the recruitment of more MRN complexes which further recruits and activates ATM (Sancar *et al.*, 2004, Blackford and Jackson, 2017). This positive feedback loop at DSBs amplifies ATM signalling to enhance activation of a DNA damage checkpoint. ɣ-H2AX functions as a docking site for many DNA damage repair factors, including 53BP1, BRCA1/2 and RAD51

(Sancar *et al.*, 2004, Blackford and Jackson, 2017). One of the other targets of ATM phosphorylation is CHK2, which targets a variety of substrates, including some that overlap with CHK1.

1.4.3) Resistance and recovery of replication forks under stress

In addition to the redundancy in the number of activated origins, another mechanism that reduces the impact of replication fork stalls is fork restart, whereby some stalled replication forks can be restarted by ATR-mediated fork repriming, upregulation of dNTP concentrations or fork remodelling (Berti and Vindigni, 2016).

Base modifications on one strand of the DNA do not block the progression of the replicative helicase, but can physically block polymerases, causing them to uncouple and expose a long track of ssDNA. Due to the dynamic nature of lagging strand synthesis, a modified base is tolerated there and causes a ssDNA lesion approximately the size of 1 or 2 Okazaki fragments (Yeeles *et al.*, 2013). However, a modified base on the leading strand is not so easily tolerated. When a fork encounters a leading strand DNA lesion, pol ε is blocked and a specialised, yet low-fidelity, polymerase can be recruited to use the lesion-containing DNA as a template, and once the lesion is bypassed pol ε can resume. The specialised polymerases are known as lesion-bypass DNA polymerases, which include pol η, κ, ι, ν, θ and REV1, 3L/7 (Sale *et al.*, 2012), and the process in which they function is known as translesion synthesis (TLS). Recruitment of lesion-bypass polymerases is facilitated by ubiquitylation of PCNA by RAD6 and RAD18. It was initially thought that leading strand DNA synthesis is continuous and could only be primed at origins. However, the ssDNA lesions seen in the replicated strand after forks encounter a lagging strand lesion can also be seen in the leading strand when forks encounter a leading strand lesion (Lopes *et al.*, 2006), implying that repriming can take place on the leading strand as an alternative to TLS. Repriming is facilitated by PrimPol, originally discovered in archaea and only more recently discovered in eukaryotes, and is the only known eukaryotic polymerase that can prime its own synthesis or prime

with DNA (Garcia-Gomez *et al.*, 2013, Mouron *et al.*, 2013). PrimPol can also function as a TLS polymerase (Mouron *et al.*, 2013). The ssDNA lesions in the replicated DNA, caused by leading or lagging strand base modifications, are repaired by homology-directed repair, or by TLS (Yeeles *et al.*, 2013, Berti and Vindigni, 2016).

TLS is categorised as a DNA damage tolerance (DDT) mechanism. Another DDT mechanism is fork reversal. When the progression of a fork is blocked on the leading strand, the newly synthesised DNA of both the leading and lagging strand can anneal, causing the fork to form a 'chicken foot' structure that resembles a Holliday junction (Neelsen and Lopes, 2015). Sub-lethal doses of topoisomerase I inhibitor do not cause DSBs in eukaryotes, previously thought to be a direct consequence of topoisomerase I inhibition, but instead cause large amounts of replication fork reversal (Ray Chaudhuri *et al.*, 2012). Additionally, combining the topoisomerase treatment with inactivation of PARP1 (poly(ADP-ribose) polymerase 1), an essential fork reversal factor, increases DNA breaks, indicating that fork reversal protects genomic stability against aberrant replication (Ray Chaudhuri *et al.*, 2012). Fork reversal is not unique to topoisomerase I inhibition, and has been identified as a global responder to replication stress in human cells (Zellweger *et al.*, 2015). Additionally, fork reversal is a physiological process that occurs at a low frequency in unperturbed and untransformed human cells (Zellweger *et al.*, 2015, Ray Chaudhuri *et al.*, 2012, Neelsen and Lopes, 2015). It is not fully understood how replication forks are reversed. *In vitro* experiments reveal that a series of DNA translocases can support fork reversal, as well as a few helicases, including RECQ family members BLM (Bloom syndrome protein) and WRN (Werner syndrome protein) (Berti and Vindigni, 2016). Also implicated in fork reversal are DSB-independent roles for recombination factors, especially RAD51 (Petermann *et al.*, 2010, Zellweger *et al.*, 2015, Berti and Vindigni, 2016). It is thought that RAD51 is recruited to stalled forks and replaces RPA at the leading strand ssDNA lesion, causing the leading strand template to re-anneal to the lagging strand template (opposite of helicase activity, reversal). The newly synthesised strands are then free to anneal, forming the Holliday junction, and RAD51 may be recruited again to

the ssDNA overhang of the newly annealed duplex, and facilitate strand invasion to the corresponding homologous sequence in front of the reversed fork (see figure 3A of Neelson and Lopes, 2015) (Zellweger *et al.*, 2015, Neelsen and Lopes, 2015). An alternative pathway to restarting reversed forks involves the helicase activity of RECQ1, as regulated by PARP1, which inhibits RECQ1 until the DNA damage that caused the fork to stall is repaired (Berti *et al.*, 2013, Berti and Vindigni, 2016). Another pathway includes the DNA2 nuclease and WRN helicase which function together to resect the annealed newly synthesised DNA, creating a 3' overhang, which is the preferential substrate for the SWI/SNF-related translocase SMARCAL1, which can convert Holliday junctions into functional replication forks (Betous *et al.*, 2013, Berti and Vindigni, 2016).

Replication forks can also be restarted after uncoupling between the CMG and polymerases without reversing first. The long track of RPA-coated ssDNA in the leading strand can be primed by pol α-primase, which requires the replication initiation factor TOPBP1 (Yeeles *et al.*, 2013), both of which promote the recruitment of the 9-1-1 complex, and thus contribute to the activation of ATR (Yan and Michael, 2009).

1.4.4) 'Double fork stalling'

Despite the measures taken to reduce the impact of fork stalling, it is still theoretically possible that two converging replication forks from neighbouring origins may both irreversibly stall, with no intervening dormant origins, a phenomenon termed a 'double fork stall' (DFS). Figure 1 is a simple schematic of how a DFS may occur. Four replication origins (W, X, Y and Z) are on a generic section of DNA, and are licensed with MCM2-7 double hexamers (blue). Upon the onset of S phase, MCMs are activated and forks progress (pink). Origins W, Y and Z are activated, and X remains dormant. The right-hand fork of origin W (W_R) and the left-hand fork of origin Y (Y_L) both stall, however, dormant origin X is stochastically activated and completes replication of its replicon. Y_R and Z_L also stalled, but in the absence of a dormant origin or fork restart, the intervening

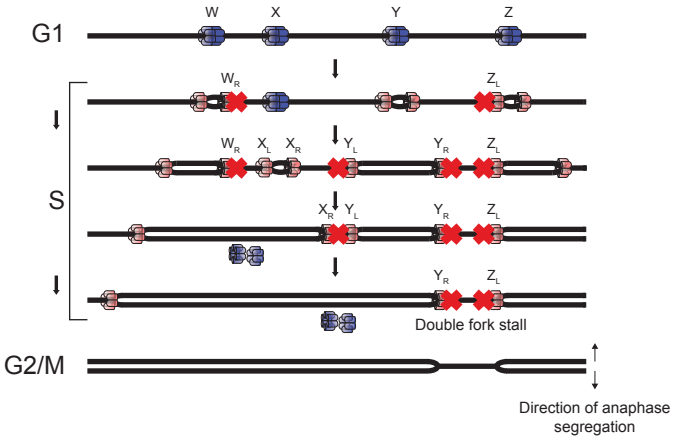


Figure 1: Double fork stalling. Cartoon depicting how replication forks that irreversibly stall can result in a double fork stall. Four replication origins have been licensed as indicated by the blue double hexamers (W, X, Y and Z). Elongating replication forks are coloured pink. Irreversible fork stalling is indicated by red crosses.

segment of DNA remains unreplicated. If left intact through to mitosis, the unreplicated DNA will physically link sister-chromatids and interfere with chromosome segregation in anaphase.

1.5) Replication origin mapping

As mentioned previously, *S. cerevisiae* origins were discovered to have an AT-rich 11-17 bp consensus sequence termed ACS (ARS consequence sequence). There are ~12000 ACSs in the ~11 Mbp haploid *S. cerevisiae* genome, but only ~500 are functional replication origins (Nieduszynski *et al.*, 2007, Siow *et al.*, 2012). ACSs require cis-acting sequence elements which include three other domains (B1, B2 and B3), together spanning a ~200 bp domain comprising the ARS. Additionally, transcription ablates origin activation, and chromatin structure can aid ORC recruitment (Nieduszynski *et al.*, 2007, Siow *et al.*, 2012). *S. cerevisiae* origins were then confirmed or dismissed by assays including 2D-gel electrophoresis, that detects intermediate replication DNA structures at specific origins, and genome-wide approaches that either detect origin-associated proteins or DNA synthesis (Nieduszynski *et al.*, 2007, Siow *et al.*, 2012).

In human cells, but also generally applicable to higher-eukaryotes, there is no origin consensus sequence as there are in yeasts. A small number of well-studied efficient human origins are restricted to specific genomic loci, but with no identifiable commonality that might function to promote pre-RC formation. More recently, genome sequencing technologies have provided comprehensive surveys of origin activation (Prioleau and MacAlpine, 2016). Origin mapping has also been attempted by chromatin immunoprecipitation of pre-RC components such as ORC and MCM. However, not all ORC binding sites are replication origins, and not all licensed origins are activated, so, although these methods may identify where origins potentially form, they are not informative of where replication forks will be activated. With respect to DFSs, active origins are more relevant so genome-wide sequencing techniques that detect active replication are summarised below.

Sequencing of short nascent strands (SNS) is designed to capture the 5' region of the leading strand, which corresponds to the DNA that is first replicated at each activated origin. In summary, this is usually achieved by heat denaturation of genomic material followed by sucrose gradient fractionation. The fraction that contains ~0.5-2.5 kbp DNA fragments is selected and digested with excess λ-exonuclease, purified and sequenced using next-generation Illumina technology. The sequenced reads are then mapped to a reference genome of the relevant cell type. Nascent strands feature a ~20 bp RNA primer at their 5' end, but includes Okazaki fragments in addition to the leading strand. Selecting for >500 bp fragments after heat denaturation will enrich leading strand fragments, and select against Okazaki fragments. λ-exonuclease is a 5' to 3' DNA exonuclease, and will digest DNA fragments without an RNA primer, and is used in excess as the enzyme has a bias against certain DNA sequences, such as G-quadruplexes, that would cause the inclusion of false positives (Foulk *et al.*, 2015). SNS sequencing is able to resolve origins to less than1000 bp (Prioleau and MacAlpine, 2016). Genome-wide SNS sequencing in a variety of human cells, including HeLa and IMR-90, identified 80,000-250,000 replication origins (Cayrou *et al.*, 2015, Besnard *et al.*, 2012, Martin *et al.*, 2011, Picard *et al.*, 2014). In one study, the origin positions were compared between multiple human cell lines and it was found that a third of the total mapped origins overlapped between cell lines (Besnard *et al.*, 2012, Picard *et al.*, 2014). These origins are at well-defined positions and are associated with CpG islands and DNase hypersensitive chromatin (open chromatin) (Cayrou *et al.*, 2015, Besnard *et al.*, 2012, Martin *et al.*, 2011). CpG islands are more commonly associated with promotion of transcription, but may have a dual role. Although a significant proportion of origins are associated with CpG islands and open chromatin, these features are neither necessary nor sufficient to define a genomic locus as an origin (Prioleau and MacAlpine, 2016).

Low percentage agarose can be used to trap replication intermediate bubble-structures (Mesner *et al.*, 2006). Prior to bubble trapping, the chromosome must be fragmented by restriction digestion, which limits the resolution of origin position to the positions of restriction sites (Prioleau and

MacAlpine, 2016). In human cell line GM06990, whole-genome bubble-seq identified ~120,000 origins, which largely occupy initiation zones, and revealed that they are enriched at DNase hypersensitive chromatin, and early origins are associated with actively transcribed genes, in partial concordance with the SNS-seq experiments (Mesner *et al.*, 2013).

At replication origins, there is a sharp transition on each individual DNA strand between the continuously (leading) and discontinuously (lagging) synthesised DNA fragments. Genome-wide mapping of Okazaki fragments in *S. cerevisiae* cells was facilitated by inactivation of DNA ligase I (Smith and Whitehouse, 2012). Replication origins largely corresponded to the transition of Okazaki fragments mapping to one strand and then the other (Smith and Whitehouse, 2012). The same technique was applied to human cells where only 66 origins could be identified by a sharp transition that occurred within a 5 kbp window. However, broad initiation zones were identified, and were on average ~30 kbp in length (Petryk *et al.*, 2016). As with SNS- and bubble-seq, the OK-seq initiation zones corresponded to CpG islands and expressed genes (Petryk *et al.*, 2016, Prioleau and MacAlpine, 2016). Additionally, OK-seq revealed that the direction of replication is co-orientated with the direction of transcription, potentially to reduce the frequency of head-to-head collision between replicative and transcriptional machinery (Petryk *et al.*, 2016, Prioleau and MacAlpine, 2016).

Collating the origin mapping data of human cells using the methods described above highlights the stochastic nature of origin positioning and activation, and how this contrasts to yeasts, where the genomic positions of most origins are known at base pair-resolution and the efficiency of many origins characterised. There is a low level of concordance in origin positions in human cells from inter- or intralab-datasets, which is likely due to the stochastic nature of human origin licensing and activation, innate biases of the different techniques used and differences between the cell lines used.

1.6) Common fragile sites and mechanisms for responding to under-replication

The first results chapter will open with theoretical approaches to predicting the frequency of DFSs in a given human cell S phase. The rest of the results presented in this thesis will first provide evidence for the validity of the model (chapter 3), and then follow-up with other cellular responses (chapter 4). Candidates for factors involved in the response to DFSs were inspired by research into CFSs (common fragile sites), which will be introduced here.

CFSs are specific genomic loci, of which there are ~90 throughout the genome of cultured human somatic cells (Durkin and Glover, 2007), where breaks occur in chromatid arms, observable in metaphase chromosomes (Glover *et al.*, 1984). These sites are hyper-sensitive to replicative stress, such as treatment with a low concentration of the replicative polymerase inhibitor aphidicolin, after which they reliably develop into breaks, which was how a large proportion of these sites were defined (Glover *et al.*, 1984). More recently, in work largely carried out by Professor Ian Hickson and colleagues, it was discovered that a subset of CFSs can become connected between sister-chromatid arms during anaphase, forming a structure termed an ultrafine anaphase bridge (Chan *et al.*, 2007) (UFB). These structures are undetectable by DAPI staining, and do not contain histones; distinguishing them from previously described 'bulky' anaphase bridges that are often lagging chromosome arms (Chan *et al.*, 2007). UFBs are detected by proteins that coat the structure: BLM helicase complex, which includes BLM helicase (Bloom's syndrome protein), topoisomerase IIIα, RMI1 and RMI2 (RecQ-mediated genome instability 1 and 2), and PICH (PLK1-interacting checkpoint helicase) (Baumann *et al.*, 2007, Chan and Hickson, 2009, Chan *et al.*, 2007). This composition of repair proteins maintains genome stability, and can facilitate branch migration in the resolution of Holliday junctions and disentangle DNA catenanes (Bachrati and Hickson, 2009, Bizard and Hickson, 2014, Chan *et al.*, 2007, Karow *et al.*, 2000, Mankouri and Hickson, 2007). BLM is a RecQ-like helicase that can unwind a variety of DNA substrates (Brosh *et al.*, 2001, Karow *et al.*, 1997, Mohaghegh *et al.*, 2001, Sun *et al.*, 1998). BLM is defective in sufferers of Bloom syndrome, who are prone to tumours, retarded growth and light sensitivity. BLM syndrome

cells exhibit an increased frequency of micronuclei (segments of chromosomes that are broken away and not segregated correctly during anaphase; their presence is a sign of genomic instability) (Chan *et al.*, 2009, German, 1993, Rosin and German, 1985). It has been proposed that the BLM complex functions to resolve segregation of genomic loci that are sensitive to replication stress (Chan *et al.*, 2007, Chan *et al.*, 2009, Liu *et al.*, 2014). Cells that lack BLM were less able to restart replication forks following treatment and release from aphidicolin (Davies *et al.*, 2007).

The frequency of UFBs increases in cells treated with low concentrations of aphidicolin (Chan *et al.*, 2009). The frequency of UFBs stained by PICH increases in a BLM-deleted cell line ,and BLM is only recruited to bridges in anaphase that are coated by PICH in metaphase (Chan *et al.*, 2007). There are two types of UFB bridges described which are distinguished by the presence or absence of a FANCD2/I foci at the termini of each bridge (Chan *et al.*, 2009). FANCD2 and FANCI are two members of nineteen in the Fanconi anaemia complex (FA complex) which has a well described role in DNA repair, especially repair of inter-strand cross-linked (ICL) DNA (Moldovan and D'Andrea, 2009). A mutation in any of these nineteen genes causes Fanconi anaemia, a disease with characteristics that overlap with Bloom's syndrome, that include cancer predisposition and retarded growth (Kee and D'Andrea, 2012). Five FA complex components (A, F, G, C and E), RPA and topoisomerase IIIα were pulled down by immunoprecipitation of BLM, and together were defined as a functional complex termed BRAFT (BLM, RPA, FA and topoisomerase IIIα) that is involved in DNA repair by DNA-unwinding (Hemphill *et al.*, 2009, Meetei *et al.*, 2003). Cells from FA patients are hyper-sensitive to reagents that cause ICLs (such as cisplatin or mitomycin-C) and also present chromosomal abnormalities that are exacerbated by replicative stress (treatment with aphidicolin) (Kee and D'Andrea, 2010, Naim and Rosselli, 2009, Schlacher *et al.*, 2012). More recently, it was discovered that depletion of FANCD2 is similar to the depletion of BLM, as well as the double-depletion, with respect to fork restart after replicative stress, indicating that they function in the same pathway to facilitate fork restart (Chaudhury *et al.*, 2013). FANCM detects stalled replication forks and recruits the FA core complex, resulting in the monoubiquitylation of FANCD2

and FANCI (Deans and West, 2009, Garcia-Higuera *et al.*, 2001, Smogorzewska *et al.*, 2007). $FANCD2^{Ub}$ localises to the stalled fork and recruits DNA repair factors FAN1 (Fanconi-associated nuclease 1) and SLX 4 (FANCP, a Holliday junction resolvase) (MacKay *et al.*, 2010, Yamamoto *et al.*, 2011). FAN1 and SLX4 facilitate recruitment of HR-mediated DNA damage repair factors BRCA2 (FANCD1), PALB2 (FANCN) and BRIP1 (FANCJ) (Wang, 2007). An alternative pathway for fork restart by FANCD2 has been discovered. The FA core complex was found to be dispensable for fork restart by FANCD2, and a monoubiquitylation-dead mutant of FANCD2 still localised to stalled forks as induced by aphidicolin and facilitated fork restart (Raghunandan *et al.*, 2015). From this it was proposed that the ubiquitylation status of FANCD2 defined the DNA damage response downstream of the FA pathway, ICL-specific repair requires $FANCD2^{Ub}$ but is dispensable for fork restart (Raghunandan *et al.*, 2015). Additionally, it was found that FAN1 is dispensable for ICL repair, but is vital for fork restart (Lachaud *et al.*, 2016).

UFBs that lack adjacent FANCD2/I foci link centromeric loci, and constitute ~55% of all UFBs. They are thought to occur because centromeres are replicated late in S phase and are the last region of a metaphase chromosome where cohesin complexes are removed (Chan *et al.*, 2007) and are proposed to contain unresolved catenanes (Wang *et al.*, 2010). UFBs flanked at each end by FAND2/I foci are products of replication stress as treatment with aphidicolin increases UFB frequency overall, and this increase is largely in the FANCD2-positive population of UFBs (Chan *et al.*, 2009, Naim and Rosselli, 2009). It was also observed that FANCD2 foci are frequently localised in pairs, from G2 through to prophase until segregating symmetrically during anaphase, and are at a similar frequency in each daughter telophase nucleus (Chan *et al.*, 2009, Naim and Rosselli, 2009). It has more recently been shown that the structure-specific endonuclease, MUS81, is required for genomic stability at CFS loci (Ying *et al.*, 2013). MUS81 requires a partner protein, EME1, to function, and was originally thought to process recombination intermediates, as *S. pombe* cells with defective MUS81 or EME1 exhibited meiotic defects (Boddy *et al.*, 2001). Since then, however, it was found that the heterodimer more efficiently (~75 x) cleaves replication fork

structures and 3'-flaps (Ciccia *et al.*, 2003), and deletion of MUS81 in yeasts results in hypersensitivity to replicative stress (treatment with hydroxyurea/camptothecin), and is proposed to have an important role in replication fork restart (Bastin-Shanower *et al.*, 2003, Doe *et al.*, 2002, Hanada *et al.*, 2007, Whitby *et al.*, 2003). MUS81-depleted human cells exhibit an increased frequency of stalled replication forks, micronuclei, bulky anaphase bridges and UFBs. Additionally, MUS81 foci colocalise with CFSs in early-mitosis as detected by FISH (fluorescence *in situ* hybridisation) (Ying *et al.*, 2013). From this, it is proposed that MUS81-EME1 cleaves incompletely replicated CFSs in early-mitosis to stop anaphase bridging and the resultant DNA break is repaired in the following cell cycle, and that this pathway is preferable to uncontrolled breakages caused by tension during chromosome segregation, or aneuploidy (Ying *et al.*, 2013).

To summarise, CFSs are sensitive to replicative stress and can develop into UFBs that link sister-chromatids. DNA repair factors are recruited to these sites before, during and after bridge formation in an attempt to conserve genomic stability, whether by fork restart (FANCD2, BLM, MUS81 etc), bridge decatenation (BLM, topoisomerase IIIα) or post hoc DNA damage repair (caused by MUS81-mediated excision).

Whether UFBs are processed by MUS81-EME1-mediated cleavage or BLM helicase unwinding, the unusual DNA structures created by this process are inherited by each daughter cell, where they are processed or repaired. In G1, there exist nuclear bodies that contain a large accumulation of a protein called 53BP1 (Harrigan *et al.*, 2011, Lukas *et al.*, 2011). They are present in unperturbed cells, but increase in frequency in response to replicative stress as caused by treatment with a low concentration of aphidicolin (Harrigan *et al.*, 2011, Lukas *et al.*, 2011). It was also observed that 53BP1 (p53-binding protein 1) nuclear bodies are frequently (~70%) symmetrical in number and morphology between sister-daughter cells (Lukas *et al.*, 2011). ChIP-qPCR revealed that 53BP1 is enriched at the FRA3B and FRA16D CFSs (the best studied and most fragile CFSs) (Lukas *et al.*, 2011). Additionally, depletion of MUS81 or BLM increases the frequency of 53BP1 nuclear bodies

in the following G1 phase (Lukas *et al.*, 2011, Ying *et al.*, 2013). Taken together, it seems possible that unreplicated DNA transmitted through mitosis results in the inheritance, by each daughter cell, of unusual DNA structures that correspond to CFSs and are coated with 53BP1.

53BP1was discovered by its binding to p53 in a yeast-2-hybrid screen (Iwabuchi *et al.*, 1994). Domain analysis revealed that 53BP1 contains tandem BRCT (BRCA1 C-terminal) domains, implicating a function in the DNA damage response (Bork *et al.*, 1997, Callebaut and Mornon, 1997). At its N-terminus are (S/T)Q motifs which were found to be hyperphosphorylated by ATM in response to DNA damaging agents, which is necessary for the rapid relocalisation of 53BP1 to damaged DNA sites (Anderson *et al.*, 2001, Rappold *et al.*, 2001, Schultz *et al.*, 2000, Xia *et al.*, 2001). As little as 30 minutes after irradiation (X-ray, infra-red or ɣ-), 53BP1 relocalises to nuclear loci that also contain previously known double strand break repair proteins ɣ-H2AX and MRN (Anderson *et al.*, 2001, Rappold *et al.*, 2001, Schultz *et al.*, 2000). 53BP1-deficient mice models are hypersensitive to irradiation and prone to tumour development (Morales *et al.*, 2003, Ward *et al.*, 2003b). 53BP1 has no known enzymatic activity, but is thought to bind to chromatin and act as a scaffold and signal transducer for other DNA repair factors (Panier and Boulton, 2014). As detailed in Panier and Boulton's comprehensive review, 53BP1 accumulates at damaged chromatin by binding to histones displaying H4K20me2 (via tandem tudor domains) and H2AK13/15Ub (via a ubiquitylation-dependent recruitment domain) modifications (Panier and Boulton, 2014). E3 ubiquitin ligases RNF8 and RNF168 prime histones for binding by 53BP1 by facilitating the ubiquitylation of L3MBTL1 and JMJD2A (marking them for deactivation or proteolysis), which mask H4K20me2, and ubiquitylation of H2AK13/15. RNF8-RNF168 activity is promoted at damaged sites by a positive feedback loop between ATM, MDC1 and MRN. DSBs are recognised by MRN which recruits autophosphorylated ATM. ATM phosphorylates histone variant H2AX to produce ɣ-H2AX, which recruits MDC1. MDC1 recruits further ATM and MRN to the damaged chromatin where activated ATM phosphorylates MDC1 which finally recruits RNF8-RNF168 (Panier and Boulton, 2014). 53BP1 binding to chromatin requires ɣ-H2AX (Ward *et al.*, 2003a) and

γ-H2AX foci span ~1 Mbp around a DSB (Rogakou *et al.*, 1999). 53BP1 functions as a key determinant of DSB repair pathway choice between homologous recombination (HR) or non-homologous end-joining (NHEJ) (Bouwman *et al.*, 2010, Bunting *et al.*, 2010, Chapman *et al.*, 2012). HR requires a homologous section of DNA elsewhere in the genome which can be used as a template for strand invasion by resected DSB ends, and facilitates high-fidelity repair even if the DNA damage has caused a significant deletion of genetic material (Heyer *et al.*, 2010). NHEJ merely rejoins broken DNA ends and can result in significant loss of genetic material (Lieber, 2010). 53BP1 is phosphorylated at DSBs by ATM which triggers the recruitment of RIF1 and PTIP which repress DNA end resection and BRCA1, and thereby HR (Bothmer *et al.*, 2010, Bunting *et al.*, 2010, Callen *et al.*, 2013, Chapman *et al.*, 2013, Feng *et al.*, 2013, Zimmermann *et al.*, 2013). BRCA1-deficient cells exhibit genetic instability associated with promoted levels of NHEJ, but this phenotype is rescued by also deleting 53BP1 (Bouwman *et al.*, 2010, Bunting *et al.*, 2010). Conversely, promotion of NHEJ by 53BP1 is inhibited by BRCA1, as loss of BRCA1 function causes RIF1 to accumulate at DSBs in S/G2 phase (NHEJ extends from G1 into the rest of interphase) and inhibits DNA end resection and thereby HR (Chapman *et al.*, 2013, Escribano-Diaz *et al.*, 2013, Feng *et al.*, 2013). Based on these experiments, it is clear that 53BP1 and BRCA1 function antagonistically to ensure that NHEJ and HR are mutually exclusive at any single DSB. The DSB repair pathway that prevails is highly dependent on cell cycle regulation, NHEJ is mostly confined to G1 and HR to S and G2. The repression of BRCA1 function, by 53BP1, is inhibited in S and G2 phase by S phase CDK activity. CDK2 phosphorylates CtIP which promotes its binding to BRCA1, and BRCA1-CtIP inhibits the interaction between 53BP1-RIF1, thereby enabling resection and HR (Escribano-Diaz *et al.*, 2013). CtIP promotes DNA end resection and is recruited to DSBs by the MRN complex (Sartori *et al.*, 2007). Cell cycle regulation of DSB repair is necessary, as HR cannot take place in G1 since there is no sister-chromatid that can be used as a template for strand invasion and high-fidelity repair. CDK2 is required for the G1/S transition and is active up until the end of G2 phase, during which period template DNA will be made available for HR by DNA replication from the beginning of S phase.

Recently, a novel form of DNA synthesis was discovered to take place in mitosis in response to replicative stress (Minocherhomji *et al.*, 2015). It was observed that punctate foci of EdU incorporation are present in prophase and prometaphase cells but only at a low frequency unless replicative stress is induced by treatment with a low concentration of aphidicolin. These foci (termed MiDAS, mitotic DNA synthesis) colocalise with FISH probes for FRA3B and FRA16D CFSs, and a majority colocalise with FANCD2 twin-foci, additionally indicating associations with CFSs (Chan *et al.*, 2009). MiDAS foci are also associated with a majority of DNA breaks in metaphase chromosome spreads. Incorporation of EdU was found to be restricted to prophase and prometaphase, as releasing cells from a nocodazole synchronisation into EdU yielded very few foci. A major event in early-mitosis is the condensation of chromatin into discretely identifiable chromosomes, a process which requires a ring-shaped complex called condensin. Depleting cells of SMC2, a member of the condensin complex, supresses chromatin condensation, as well as MUS81 recruitment to CFSs, and EdU incorporation at CFSs in early-mitotic cells. Depletion of MUS81 nuclease activity reduces the recruitment of POLD3 (a non-catalytic subunit of DNA pol δ) to chromatin, and reduces the frequency of MiDAS foci. Depletion of POLD3, but not POLD4 (a catalytic subunit of DNA polymerase δ), reduces the frequency of MiDAS foci. In human cancer cells synchronised at the G1/S boundary, POLD3 is dispensable for canonical DNA synthesis and fork speed is unaffected, although cells progress through S phase slowly (Tumini *et al.*, 2016), however, POLD3 is haploinsufficient in mice (Murga *et al.*, 2016). POLD3 is the human orthologue to *S. cerevisiae* Pol32, which is required for BIR (break-induced replication) and ALT (alternative lengthening of telomeres; telomere elongation in the absence of telomerase) (Lydeard *et al.*, 2007). BIR is a repair pathway initiated by a single recombination strand invasion event, followed by up to 100 kbp of DNA synthesis (Malkova *et al.*, 2005), and reinitiates DNA synthesis at damaged replication forks and repairs one-ended DSBs in yeasts (McEachern and Haber, 2006), by extension of a migrating D-loop (Donnianni and Symington, 2013). BIR requires CMG (CDC45, MCM2-7, GINS) as well as CDC7 and CDT1, but not ORC or CDC6 (Lydeard *et al.*, 2010) and facilitate conservative replication (Donnianni and Symington, 2013, Saini *et al.*, 2013). Minocherhomji *et al.*

2015, also discovered that suppression of MiDAS by aphidicolin increased non-disjunction, UFBs, DNA breaks in metaphase chromosome spreads, 53BP1 nuclear bodies in the following G1 and reduced cell survival. Taken together, the authors propose a model whereby stalled replication forks at CFSs are excised by MUS81-EME1 nuclease to create a single-ended DSB which is resected for strand invasion. A D-loop is formed and migrates with BIR-like POLD3-dependant conservative replication, which reinitiates replication at that locus. MUS81 recruitment requires chromosomal condensation, so the authors speculate whether the mechanical forces of condensation expose unreplicated DNA as a template for DNA repair synthesis. Successful gap filling by DNA repair synthesis reduces the amount of unreplicated DNA transmitted through anaphase. Follow up work from the same group found that RAD52 is required for MiDAS foci formation, independently of ATR, RAD51 and BRCA2 (Bhowmick *et al.*, 2016), implicating a novel role for RAD52 which so far has been thought to be backup in RAD51-depleted cells. The authors propose that RAD52 plays a specialised role in mitosis to facilitate microhomology-mediated BIR, a class of BIR which can occur independently of RAD51, by annealing a small overhang at a one-ended DSB (created by MUS81-EME1 nucleolytic cleavage at a stalled fork) to a microhomology in the unreplicated DNA. From there, POLD3-dependant DNA synthesis conservatively replicates the previously unreplicated DNA (Bhowmick *et al.*, 2016).

A number of the experiments carried out on MiDAS have used aphidicolin to cause CFS expression, and the CDK1-inhibitor RO-3306 to synchronise cells at the G2/M boundary, in an attempt to ensure that the DNA synthesis they observe takes place in mitosis, and not in G2. In human cells, CDK1 primarily regulates the G2-M phase transition. RO-3306, a small molecule inhibitor of CDK1, functions as an ATP-competitor at the ATP binding pocket of CDK1 (Vassilev *et al.*, 2006). CDK1 and CDK2 are highly structurally similar, so inhibitors of CDK1 also inhibit CDK2, but RO-3306 has, so far, the greatest selectivity for CDK1 over CDK2 (~10 fold) (Vassilev, 2006, Vassilev *et al.*, 2006). CDK2 promotes the G1/S phase transition, as well as progression through S phase. Although RO-3306 is described as a strong synchroniser of cells at the G2/M

boundary, which can then be efficiently released into mitosis, the synchronisation is not absolute. Minocherhomji *et al.* 2015 show in figure S2C the dynamics of RO-3306 release in U2OS cells, and see that 5 minutes after release ~one third of cells are still in G2 and 20 mins after release 10% of cells are in G2. The authors incubate U2OS cells in EdU for 30 mins after release from RO-3306, so one sixth of the incubation time takes place when between ~33-100% of cells are in G2 and one half of EdU incorporation takes place when 10-33% of cells are in G2. It is likely that a majority of the MiDAS foci the authors observed were *de novo* DNA synthesis in mitosis, however, it is also possible that a significant proportion of MiDAS foci are from late replication in G2 (estimated to be ~1% of the genome) (Widrow *et al.*, 1998) or DNA repair synthesis including fork restart in G2, translesion synthesis (Diamant *et al.*, 2012) or homologous recombination. Additionally, combining a polymerase inhibitor (aphidicolin) with a molecule that also inhibits CDK2 activity is highly unphysiological.

Taken together, it is proposed that FANCD2 and MUS81 are recruited to stalled replication forks that are associated with unreplicated segments of DNA, and facilitate unlinking of sister-chromatids before anaphase. If unsuccessful, UFBs form and either BLMs helicase facilitates unwinding or mis-segregation events take place. Whether MUS81 endonuclease activity excises unreplicated DNA, or BLM helicase unwinds unreplicated DNA into stretches of ssDNA, the resultant DNA damage is largely inherited symmetrically by each daughter cell (Lukas *et al.*, 2011). It has also been proposed that, in an alternative process, that MUS81 and FANCD2 facilitate POLD3- and RAD52-dependent BIR at unreplicated segments of DNA in early-mitosis (Bhowmick *et al.*, 2016, Minocherhomji *et al.*, 2015). Depletion or repression of BLM, FANCD2, MUS81, MiDAS, POLD3 and RAD52 increase the frequency of G1-specific 53BP1 nuclear bodies (Bhowmick *et al.*, 2016, Bourseguin *et al.*, 2016, Lukas *et al.*, 2011, Minocherhomji *et al.*, 2015, Ying *et al.*, 2013). Lukas *et al.* 2011 proposed that CFS breaks are caused by tension from chromatin condensation in early-mitosis, based on the increase in UFB frequency and the reduction in 53BP1 nuclear body frequency caused by SMC2 depletion, indicating that condensation caused the breaks that 53BP1

eventually coats, and if not, would form a UFB as a safer alternative. However, Ying *et al.* 2013 observe that depletion of MUS81 reduces the frequency of chromosome breaks but an increase in UFBs and segregation defects (bulky bridges, micronuclei), and counter-propose that breaks seen in early-mitosis are caused by MUS81 nuclease, not by mechanical forces from condensation, and are preferable to forming UFBs. Additionally, in figure S5B and S5C, Minocherhomji *et al.* 2015 show that 53BP1 nuclear bodies increase in frequency after SMC2 depletion. Either way, both possibilities are likely to result in DSBs. Alternatively, it is proposed that unreplicated DNA in UFBs is unwound by helicase activity, considering the presence of BLM helicase and topoisomerase IIIα at UFBs, which together could facilitate decatenation of hemicatenated DNA, and find in support of this the presence of BrdU after partial DNA denaturation, indicating ssDNA, across the length of UFBs (Chan *et al.*, 2007), as well as RPA (Chan and Hickson, 2009). However, the absence of detectable ssDNA in G1-specific 53BP1 nuclear bodies, whether by BrdU in non-denaturing conditions or RPA (Harrigan *et al.*, 2011, Lukas *et al.*, 2011), throws doubt on this possibility.

Research into CFSs is a good foundation for potential DFS responses. They are considered origin-poor regions. A DNA fibre study of forks at FRA3B and FRA16D (the two most studied and highly expressed CFSs) showed that the fork speed and fork stall rate were the same in the CFSs as elsewhere in the genome, but found at FRA3B a ~700 kbp region of DNA that contained no initiation events, with consistent results at FRA16D (Letessier *et al.*, 2011) and FRA6E (Palumbo *et al.*, 2010). It is this category of CFSs that are expected to overlap with genomic loci that are prone to spontaneous DFSs, as opposed to CFSs that are expressed at large genes (making a collision with transcription factors probable). A large amount of CFS research has been based on colocalisation experiments specifically between FRA3B and FRA16D (as probed by FISH), and DNA damage responders including FANCD2, UFBs, MUS81 and 53BP1 (Chan *et al.*, 2009, Lukas *et al.*, 2011, Ying *et al.*, 2013).

1.7) Aims of this thesis

In collaboration with the Blow lab, the DFS model published in Newman *et al.* 2013 has been applied to origin mapping data in human cells by the same group (Al Mamun *et al.*, 2016, Moreno *et al.*, 2016).

The first aim is to find a biological marker for DFSs, or their consequences, and use it to emulate precise predictions made by the DFS model, which is presented in chapter 3.

The second aim is to use research into CFSs as a foundation to identify further candidates of the DFS response, which is presented in chapter 4.

Chapter 2

Materials and methods

2.1) Tissue culture & cell lines

Human cell lines HeLa (ATCC CCL-2), U2OS (ATCC HTB-96) and IMR-90 (ATCC CCL-186) were obtained from the American Type Culture Collection (ATCC) and used for experiments at no more than 20 population doublings. All cells were cultured in Dulbecco's Modified Eagle's Medium (DMEM; 41966, Gibco) supplemented with 10% v/v fetal bovine serum (FBS; 10270106, Gibco), 100 Units/ml penicillin and 100 µg/ml streptomycin (15140, Gibco). Cultures were maintained in incubators set to 37 °C, 5% CO_2 and in a humid environment. For culture maintenance and assay preparation, all cell lines were washed with Dulbecco's phosphate buffered saline (DPBS; 14190, Gibco) and suspended with 0.25% Trypsin-EDTA (25200, Gibco). Where appropriate, cells were counted by trypsinising culture dishes of cells and mixing a small aliquot of suspended cells in DMEM with 0.2% v/v trypan blue (T10282, Life Technologies) and transferring them to a Countess Chamber Slide (C10283, Invitrogen). Alive cells, that excluded the dye, were automatically counted by a Countess II Automated Cell Counter (AMQAX1000, Life Technologies). HeLa cells were split 1:~10, ~3 times per week, meaning that their doubling time was ~16 hours.

2.2) Cell synchronisation

2.2.1) Release into S phase

HeLa and U2OS cells were synchronised in early S phase by double thymidine block. Cells were incubated in 2.5 mM thymidine (T1895-10G, Sigma-Aldrich) in DMEM for 16 hours and then released by DPBS washes followed by an incubation in DMEM for 8 hours. Cells were then incubated in 2.5 mM thymidine for 16 hours. Finally, DPBS washes released cells into S phase.

2.2.2) Release into G1

HeLa cells were synchronised by a double thymidine block. They were released for 4 hours by DPBS washes and then treated with 50 ng/ml nocodazole (M1404, Sigma-Aldrich) for 4 hours. To release cells from metaphase, they were washed with DPBS.

2.3) RNAi & siRNAs

Transfection of short interfering RNA (siRNA) was carried out using the Lipofectamine RNAiMAX system in accordance with the manufacturer's instructions. Two mixtures were prepared in parallel; siRNA solution in Opti-MEM (31985, Gibco) and Lipofectamine RNAiMAX (137780, Invitrogen) in Opti-MEM, which were combined before addition to DMEM (with 10% FBS but without antibiotics) on cultured cells. Lyophilised siRNAs were solubilised in 5x siRNA Buffer (B-002000-UB, Dharmacon) diluted to 1x in Molecular Grade RNase-free water (B-003000-WB, Dharmacon). Transfection efficiency using this system is ~90% as measured by transfection of siGLO Green Transfection Indicator (D-001630-01, Dharmacon) using the manufacturers instructions.

Table 1: siRNA sequences and the concentrations at which they were used.

Target transcript	Sequence (starting from 5' sense)	Concentration	Product code
Control	UAGCGACUAAACACAUCAA	-	D-001210-01-50
53BP1	GAAGGACGGAGUACUAAUAUU	20 nM	Custom order
CDT1 SMARTpool	CCAAGGAGGCACAGAAGCA GCUUCAACGUGGAUGAAGU UCUCCGGGCCAGAAGAUAA GGACAUGAUGCGUAGGCGU	20 nM	M-003248-02
MCM5	GGAUCUGGCCAGCUUUGAUUU	50 nM	Custom order

All from Dharmacon.

Given the abundance and slow turnover of MCMs, a regime of transfections was developed to significantly deplete the cellular pool of MCM5. HeLa cells were transfected twice consecutively over a 32 hour period, such that the cells in the final sample will have been in transfection reagent for a total of 64 hours. During this regime, samples were harvested at 16 hour intervals starting after the first transfection, such that a variety of depletions of MCM5 can be quantified. U2OS cells were transfected for a 24 hour period twice consecutively, and so were transfected for a total of 48 hours.

2.4) Immunoblotting

2.4.1) Sample preparation

Preparation for immunoblotting was the same for each cell line. For an unfractionated 'whole cell' lysate, cell pellets were lysed by ice-cold incubation with radioimmunoprecipitation assay buffer (RIPA, 50 mM Tris-HCl pH 7.4, 150 mM NaCl, 1% v/v IGE-pal CA-630, 0.5% w/v Na-deoxycholate, 1mM EDTA, supplemented with 1 mM PMSF, 1 mM Na-OV, 10 μg/ml leupeptin, 10 μg/ml aprotinin and 10 μg/ml pepstatin). Lysates were sonicated and centrifuged to break-up and clear genomic material. The protein concentration of each lysate was quantified by the Bradford assay. Lysates were mixed with Bio-Rad Protein Assay Dye Reagent (5000006, Bio-Rad) and absorbance measured at 595 nm wavelength with FLUOstar Omega plate reader (BMG Labtech). 10 μg of protein was heated to 70 °C with NuPAGE LDS Sample Buffer (NP0007, Novex) supplemented with 25 mM dithiothreitol (DTT, 443853B, VWR) before loading into a NuPAGE 4-12% Bis-Tris Protein Gel (NP0322, Invitrogen). Samples were run alongside PageRuler Plus Prestained Protein Ladder, 10-250 kbp (26619, Thermo Scientific). Gels were run at 80 V for 30 minutes followed by 200 V for 30 minutes. Proteins were transferred onto nitrocellulose membrane, 0.2 μm pore size (10600001, Amersham) in transfer buffer (12.5 mM tris, 96 mM glycine, 10% v/v ethanol) at 40 V for 15 hours. Membranes were stained with Ponceau S Stain (K793-500ML, AMRESCO), which was used with the protein ladder to dissect the membrane. Ponceau stain was washed away with tris-buffered saline (TBS, 50 mM tris, 150 mM NaCl, pH 7.6) and membranes were blocked in 5% powdered milk w/v (Original dried skimmed-milk, Marvel) in TBS + 0.1% v/v Tween 20 (TBST, BP337-500, Fisher Scientific) for 1 hour. Primary antibodies (detailed in table 2) were diluted in 3% bovine serum albumin (BSA, A9647-100G, Sigma) and 0.01% v/v Na-azide (786-299, G-Biosciences) in TBST and membrane was incubated in this mixture overnight (~15 hours) at 4 °C. The membrane was washed in TBST followed by incubation in secondary antibodies (detailed in table 2) diluted in block solution for 1 hour at room temperature. The blots were washed with TBST and then transitioned into TBS. Fluorescence intensity was acquired with an Odyssey

CLx Imaging System (LI-COR) at 700 nm and 800 nm excitation wavelengths at a resolution of 84 µm pixels.

To isolate the chromatin-bound fraction of proteins, pellets of cells were incubated in ice-cold CSK extraction buffer (10 mM HEPES pH 7.4, 300 mM sucrose, 100 mM NaCl, 3 mM $MgCl_2$ supplemented with 1 mM PMSF, 1 mM Na-OV, 10 µg/ml leupeptin, 10 µg/ml aprotinin and 10 µg/ml pepstatin) with 0.5% v/v Triton X-100 (BP151-500, Fisher Scientific) for 10 minutes on ice. Extracted nuclei were centrifuged and the pellets washed with ice-cold CSK extraction buffer (without Triton) twice. For comparison between the soluble and chromatin-bound fractions of 53BP1, the CSK + Triton fraction was retained and loaded proportionally to the chromatin-bound fraction, otherwise they were discarded. The final pellet of extracted chromatin (and material from the nuclear lamina) was then prepared for immunoblotting, as described above.

Table 2: primary and secondary antibodies used for immunoblotting.

Antibody	Source	Manufacturer	Product code	Dilution
53BP1	Rabbit - polyclonal	Bethyl	A300-272A	1:5000
α-tubulin	Mouse - monoclonal	Sigma-Aldrich	T6199	1:20000
GAPDH	Mouse - monoclonal	Abcam	ab9484	1:1000
Lamin B1	Rabbit - polyclonal	Abcam	ab16048	1:1000
MCM5	Mouse - monoclonal	Santa Cruz	sc-136366	1:1000
IRDye 680 anti-mouse	Donkey - polyclonal	LI-COR	926-68072	1:15000
IRDye 680 anti-rabbit	Donkey - polyclonal	LI-COR	926-68073	1:15000
IRDye 800 anti-mouse	Donkey - polyclonal	LI-COR	926-32212	1:15000
IRDye 800 anti-rabbit	Donkey - polyclonal	LI-COR	926-32213	1:15000

2.4.2) Sample acquisition & analysis

Quantification of band intensity was carried out with Image Studio software (LI-COR). Background was automatically detected from the signal surrounding each region of interest (ROI, effectively each band) and subtracted per-pixel. Intensity was then normalised against the loading control. Finally, relative intensity was normalised against the negative controls, which were assigned an arbitrary value of 1.

2.5) Flow cytometry

2.5.1) Sample preparation

Preparation for flow cytometry was the same for each cell line. Immediately before harvesting cells they were incubated with 40 μM 5-ethynyl-2'-deoxyuridine (EdU, C10634, Invitrogen) added directly to the DMEM for 30 minutes. Pellets of live cells were incubated in ice-cold HEPES-CSK buffer (1% w/v BSA, 1mM EGTA, 0.2% v/v Triton X-100, 1mM DTT, 10 mM HEPES pH 7.4, 300 mM sucrose, 100 mM NaCl, 3 mM $MgCl_2$ supplemented with 1 mM PMSF, 1 mM Na-OV, 10 μg/ml leupeptin, 10 μg/ml aprotinin and 10 μg/ml pepstatin) for 10 minutes on ice. Extracted nuclei were washed with 1% w/v BSA in DPBS (14190-250, Gibco) and then fixed with 2% v/v formaldehyde (28908, Thermo Scientific) in DPBS for 15 minutes at 37 °C and washed again. Cells were permeabilised with ice-cold 70% v/v ethanol in DPBS at -20 °C for 10 minutes and then washed as above. Cells were then incubated in anti-BM28 antibody (mouse monoclonal, 1:500, 610701, BD Biosciences) in 1% BSA w/v DPBS for 2 hours and then washed. Cells were incubated with Alexa Fluor 488 anti-mouse secondary antibody (1:1000, A11029, Molecular Probes) diluted in 1% BSA w/v in DPBS for 1 hour and then washed. To conjugate a fluorophore to the incorporated EdU, a Click-iT reaction (C10634, Invitrogen) cocktail was prepared by mixing 219 μl DPBS, 5 μl 100 mM $Cu(II)SO_4$, 1 μl 1mM Alexa Fluor 647 picolyl azide and 25 μl 'Reaction Buffer Additive' (~5 mM ascorbic acid) for a final volume of 250 μl per sample. This reaction mixture was made fresh and immediately added to cells for 30 minutes, followed by a wash with 1% BSA w/v in DPBS. Finally, cells were stained with 50 μg/ml propidium iodide (PI, P4864-10ML, Sigma-Aldrich) in 1% BSA w/v in DPBS and 50 μg/ml RNase A (EN0531, Thermo Fisher).

2.5.2) Sample acquisition & analysis

Events were recorded using a BD FACS Canto II flow cytometer (BD Biosciences) with FACSDIVA software (BD Biosciences). Six parameters were recorded: forward scatter (peak area), side scatter (peak area), PI (peak width and area), Alexa Fluor 488 (peak area) and Alexa Fluor 647 (peak area). Detector voltages were adjusted for each parameter with the negative control for each experiment. Additionally, events with a chromosome content of 2n (diploid G0/G1 cells) were adjusted to 50,000 arbitrary intensity units on the PI-area scale, thus events with a chromosome content of 4n (G2/M cells) would have an arbitrary intensity of ~100,000 units.

Flow cytometry data was analysed using FlowJo software (FlowJo, LLC). Plotting events using the forward scatter and side scatter values provided a qualitative diagnostic of the quality of the cells and sample preparation. Populations of apoptotic cells, cell debris or general debris (undissolved BSA, for example) as well as doublets (cells that have clumped together during sample preparation) can be identified in this plot. Plotting events using PI peak width (PI-W) against PI peak area (PI-A) allowed the identification of cycling cells separate from apoptotic cells, doublets and debris, which were all gated out for all further analysis. To analyse chromatin-bound MCM2 immunofluorescence in early S phase cells, early S phase cells were defined as events that had above-background levels of EdU fluorescence (where background was defined by cells that had no EdU incorporation but were treated with the Click-iT reaction) but had not significantly increased their total DNA content beyond 2n (50,000 arbitrary units of PI fluorescence intensity. To quantify the average intensity of chromatin-bound MCM2 in early S phase cells, the Alexa Fluor 488 intensity values (MCM2 on chromatin) in all events that were defined as early S phase were averaged by calculating the geometric mean. Using these mean values background was removed by subtracting the mean intensity of Alexa Fluor 488 in cells that were not incubated with anti-BM28 primary antibody. After correcting for background, abundance of MCM2 on chromatin in early S phase cells was quantified relative to cells transfected with control siRNA which were assigned the arbitrary value of 100%.

2.6) Microscopy

2.6.1) Sample preparation

Sample preparation for microscopy was the same for each cell line. Thickness #1.5 coverslips (631-0156, VWR) were sterilised in 70% v/v ethanol before being placed into empty culture dishes. Suspended cells were aliquoted and incubated in these dishes until fully adhered, and seeded at a density such that prior to fixing the cells they were at ~80% confluency. For experiments requiring fluorescent staining of nucleotide incorporation, cells were incubated with 40 μM EdU for 30 minutes up until fixation. To fix cells, the coverslips were removed from culture dishes and washed with DPBS, followed by incubation in 4% v/v formaldehyde in DPBS (28908, Thermo Scientific) for 10 minutes. After washing with DPBS, cells were permeabilised with 0.2% v/v Triton X-100 in DPBS for 10 minutes. Cells were washed with 0.1% v/v Tween in DPBS (PBST) and blocked in 0.5% v/v fish skin gelatine (G-7765, Sigma) in PBST (FSG-PBST) for 1 hour. Primary antibodies (detailed in table 3) were diluted in FSG-PBST, and cells were incubated in antibody at 4 °C overnight (~15 hours). Following washes with PBST, cells were incubated in secondary antibodies (indicated in table 3) also diluted in FSG-PBST for 1 hour. Cells were then washed with PBST which was transitioned into PBS. If cells were incubated with EdU then the Click-iT reaction mixture was prepared as follows: 215 μl DPBS, 10 μl 100 mM Cu(II)SO_4, 1 μl 1mM Alexa Fluor 555 picolyl azide and 25 μl 'Reaction Buffer Additive' (~5 mM ascorbic acid) for a final volume of 250 μl per sample (C10338, Invitrogen). Cells were treated with this mixture for 30 minutes and then washed with DPBS. Finally, cells were incubated in 0.1 μg/ml 4',6-diamidino-2-phenylindole (DAPI, D9542, Sigma) in DPBS for 5 minutes and then washed with DPBS. Coverslips were mounted onto slides (Superfrost, Thermo Scientific) in VectaShield (H-1000, Vector Laboratories) and sealed with nail varnish.

Table 3: primary and secondary antibodies used for immunofluorescence microscopy.

Antibody	Source	Manufacturer	Product code	Dilution
53BP1	Rabbit - polyclonal	Novus	NB100-904	1:2000
γ-H2AX	Rabbit - polyclonal	Cell Signalling	2577	1:1000
CyclinA	Mouse - monoclonal	Abcam	ab16726	1:200
FANCD2	Rabbit - polyclonal	Santa Cruz	Sc-28194	1:200
p-H3	Rabbit – polyclonal	Cell Signalling	9701	1:1000
RPA	Mouse - monoclonal	Abcam	ab2175	1:300

The cells used to quantify the proportion of 53BP1 in nuclear bodies were a gift from Professor Jiri Lukas and express a GFP-53BP1 fusion protein, with more details in Bekker-Jensen *et al.* 2005.

2.6.2) Image acquisition

Images were acquired with a DeltaVision Elite (GE Healthcare) widefield fluorescence microscope, using a Coolsnap HQ (Roper Scientific) CCD camera and solid-state illumination. Most experiments were imaged with a 40x/1.3 Oil Plan Fluor APO objective (Olympus), except when imaging EdU foci in mitotic cells and FANCD2, where a 100x/1.4 Oil Super-Plan APO objective (Olympus) was used. The entire camera sensor chip was used for each image (1024x1024) and the exposure time and transmission were adjusted for each experiment to ensure that the most intense pixel of fluorescence was detected in the linear range of the sensor. Excitation light was filtered with the following filter set: DAPI, FITC, TRITC and Cy5, which correspond to fluorescent staining with DAPI, Alexa Fluor 488, Alexa Fluor 555 and Alexa Fluor 647 respectively. All experiments required imaging of the cell nuclei only and so required a z-section range of 10-12 μm, with sections taken every 1 μm. Z-sections were deconvolved using softWoRx Constrained Iterative Deconvolution with the Enhanced Ratio protocol for 10 cycles.

2.6.3) Image analysis

Image analysis was carried out with Volocity (Perkin Elmer). Foci were identified with a minimally-biased approach using the following sequential programme consisting of Volocity analytical functions: 1) Find Objects Using Signal Intensity (threshold set to intranuclear background), 2) Clip Objects to ROI (ROI is drawn around nucleus using DAPI signal), 3) Remove

Noise From Objects (Median Filter), 4) Separate Touching Objects (Minimum Object Size Guide = 1 μm^3), 5) Exclude Objects by Size (Exclude Objects < 0.1 μm^3). In practice, the protocol finds signal intensity above background in a specified channel, separates objects that may be overlapping and excludes small objects, all bound within the volume defined by DAPI signal.

Colocalisation analysis was carried out using the built-in colocalisation tools in Volocity. The two channels of interest were selected and each threshold was set to their respective intranuclear backgrounds. ROIs were defined by the volume occupied by DAPI signal, and within each ROI (each nucleus) Pearson's Correlation Coefficient and Manders' Overlap Coefficient was calculated. Additionally, colocalisation was calculated by a third method where foci of two separate channels (for example, 'A' and 'B') were identified as foci described above and categorised as either foci 'A' only, foci 'B' only or co-localised if foci from both channels overlapped by a single voxel or more.

2.7) Clonogenic assay

HeLa cells were transfected with siRNAs targeted against MCM5 and 53BP1 transcripts and a scramble control, as described above. Following transfection, cells were seeded into culture dishes at a density of 500 cells per 100 mm diameter petri dish. After allowing the cells to adhere for 24 hours, they were incubated in hydroxurea (HU, H8627-5G, Sigma-Aldrich) at the indicated concentrations in DMEM for 48 hours. After washing away HU with DPBS and adding fresh DMEM, colonies were allowed to grow for 10-12 days. Colonies were fixed with ice-cold 100% methanol for 20 mins at -20 °C. Colonies were stained with crystal violet solution (25% v/v methanol, 0.5% v/v crystal violet (HT90132, Sigma-Aldrich) in water) for 20 minutes and then washed with water and counted. Colony counts were normalised against the untreated counts in each transfection set, which were assigned an arbitrary value of 1.

2.8) Mathematical & computer simulated modelling

All theoretical modelling was carried out by Dr Mohammed Al Mamun, Dr Luca Albergante and Professor Timothy Newman. This work is detailed in Moreno *et al.* 2016 and Al Mamun *et al.* 2016 and the concept first published in Newman *et al.* 2013.

2.9) CDC6 overexpression HBEC cell line

This cell line was developed by Eirini-Stavroula Komseli and Professor Vassilis Gorgoulis as detailed in Petrakis *et al.* 2016. The experiments using this cell line presented in this thesis were carried out by Dr Alberto Moreno. Details of the cell line can be found in Moreno *et al.* 2016.

2.10) 53BP1 ChIP sequencing

The ChIP seq (chromatin immunoprecipitation sequencing) experiment presented in this thesis was carried out by Dr Alberto Moreno. The data was analysed by Dr Luca Albergante and Dr Mohammed Al Mamun. Details of both procedures can be found in Moreno *et al.* 2016.

2.11) Immunofluorescence staining of ultrafine anaphase bridges

Ultrafine anaphase bridges (UFBs) were stained and quantified by Dr Alberto Moreno using the anti-BLM primary antibody (sc-7790, Santa Cruz). More details can be found in Moreno *et al.* 2016.

Chapter 3

Double fork stalls frequently occur in unperturbed cells, and are marked by G1-specific 53BP1 nuclear bodies in the following cell cycle

3.1) Predicting the Frequency of Double Fork Stalls in Unperturbed Human Cells

There has long been a low level of consistency in human cell origin mapping datasets. Origin mapping of the same cells lines would produce highly variable estimates for the number and distributions of replication origins. Bubble-seq performed with the human GM06990 cell line proved to be a relatively robust and experimentally independent method to validate SNS-seq datasets (Mesner *et al.*, 2013). The authors compared their dataset with a separate groups SNS-seq dataset of the same cell line and found that ~45% of the SNS calls fell within bubble-containing EcoR1 fragments (the unit of resolution in bubble-seq analyses) (Mesner *et al.*, 2013). Picard *et al.* 2014 used these two sequencing technologies in parallel on several human cell lines, including HeLa and IMR-90 cells, and found that ~65% of SNS calls were validated by bubble-seq (expected to be ~6% with randomised datasets) (Picard *et al.*, 2014). Analysis of the Picard dataset finds that there are ~180,000 replication origins in the diploid human genome of ~5.65 Gbp, which means that the mean replicon size is ~31 kbp in both HeLa and IMR-90 cells (Al Mamun *et al.*, 2016, Picard *et al.*, 2014). In any individual unperturbed human cell, it is estimated that ~30% of origins are activated in a S phase (Ge and Blow, 2010, Ge *et al.*, 2007, Ibarra *et al.*, 2008, Kunnev *et al.*, 2015), and DNA fibre analysis shows that origin activations are an average of ~100 kbp apart (Conti *et al.*, 2007, Jackson and Pombo, 1998). So, if ~30% of origins are ~100 kbp apart in an individual cell, then all origins within a population of cells are estimated to be an average of ~33 kbp apart, which is consistent with the frequency of origins detected by Picard *et al.* 2014. Although the total number of origins seem to be a good fit to independent experiments, it should be noted that the largest replicons within a dataset has a disproportionate influence on the predicted frequency of DFSs (Al Mamun *et al.*, 2016), and it is possible that a single false-negative origin could as much as half the size of the largest replicon.

The theoretical model, published by Newman *et al.* 2013, that predicts the probability of at least one DFS event in a given eukaryotic S phase was developed with the assumption that the largest replicon is significantly smaller than the estimated median number of base-pairs a replication fork

replicates before irreversibly stalling (Al Mamun *et al.*, 2016, Newman *et al.*, 2013). The *S. cerevisiae* genome conformed to this constraint as the largest replicon is 60 kbp and the median stall distance is ~10 Mbp (Newman *et al.*, 2013, Siow *et al.*, 2012). However, analysis of the HeLa origin mapping data published by Picard *et al.* 2014 revealed that the distribution of replication origins is highly irregular compared to yeasts (Al Mamun *et al.*, 2016). A measure of regularity is given by the coefficient of variation (*R*, standard deviation/mean) which for absolutely equal replicon sizes is 0 and for random replicon lengths is 1. Replicon lengths in the *S. cerevisiae* genome have an *R*-value of ~0.7 (more uniform than random) and in the HeLa genome it's ~1.5 (more irregular than randomly distributed origins) (Al Mamun *et al.*, 2016, Newman *et al.*, 2013, Picard *et al.*, 2014). This irregularity is also illustrated in figure 2A, the frequency distribution of HeLa replicons, where the mean replicon length (red) of ~31 kbp is more than two orders of magnitude smaller than the largest replicons (Al Mamun *et al.*, 2016, Picard *et al.*, 2014). In fact, the HeLa genome contains an unexpectedly high frequency of very large replicons, the largest of which is ~6 Mbp (which would be ~0.3 Mbp if origins are distributed randomly) (Al Mamun *et al.*, 2016, Picard *et al.*, 2014).

Given the presence of very large replicons in the HeLa origin mapping dataset, which approach the expected median stall distance, a general model, termed the "central equation", was developed from the same principals used to develop the DFS model published by Newman *et al.* 2013 (Al Mamun *et al.*, 2016) (figure 2B). The central equation does not contain the approximations that were used in the previous DFS model and so appears more complicated. However, both models share the same key parameters each of which have similar properties. For example, inputting the parameters of *S. cerevisiae* origin-mapping data into both versions of the model yields the same predicted probability of DFSs of ~0.1% (Al Mamun *et al.*, 2016, Newman *et al.*, 2013, Siow *et al.*, 2012). The key parameters in both models are the genome size (N_g), median stall distance (N_s), median replicon length/origin number (N_l or K) and the origin distribution (R). The average replicon length is equivalent to origin number within a dataset: $N_l \approx N_g/(K+1)$. The coefficient of variation, R, is

reflected by the sum of N_i (length of the i^{th} origin) over K origins, for example, huge replicons will increase the standard deviation (and thereby R) and reduce the magnitude of the summation on the right-hand side of the central equation (figure 2B), both of which result in increased predicted probabilities for DFSs. The only poorly-known parameter is the average number of bases replicated by a fork before it irreversibly stalls, N_s (median stall distance). However, Newman *et al.* 2013 used a variety of approaches to estimate the value of N_s, including by quantifying the frequency of fork stalling in DNA fibre assays published by Maya-Mendoza *et al.* 2007. A particular pattern in the fluorescently labelled incorporated nucleotides can be identified as fork stalls and occur in ~0.16% of DNA tracks, where each track is ~25 kbp. Therefore, the per nucleotide stall rate can be calculated as 0.0016/25000 bp to be ~6 x 10^{-8} bp^{-1}. Using an intermediate equation in the development of the earlier DFS model, Newman *et al.* 2013 derived N_s from the per nucleotide stall rate to be ~10 Mbp ($N_s \approx \ln(2)/ \sim 6 \times 10^{-8}$ bp^{-1}), but stressed that this should be considered a rough estimate. The DNA fibre assays were carried out with MRC5 cells, but the median stall rate of ~10 Mbp was found to be well conserved with the estimated *S. cerevisiae* fork stall rate derived by theoretical methods. Newman *et al.* 2013 estimated the theoretical fork stall rate by making the assumption that the frequency of DFS at telomeres should be approximately equal to the rest of the genome. These regions are considered separately in the theoretical model as a single fork stall in a telomeric region is equivalent to a DFS in a replicon. The probability of single fork stalls at all *S. cerevisiae* telomeres is therefore modelled as a function of N_s, as opposed to N_s^2 in replicons. Assuming the probability of single fork stalls at telomeres to be similar to DFSs in all replicons means that the equation can be rearranged to solve for N_s (see equation 9 in Newman *et al.* 2013). Inputting the parameters from the *S. cerevisiae* origin mapping data into this equation outputs an estimate for N_s as ~12.7 Mbp (per nucleotide stall rate of 5.4 x 10^{-8} bp^{-1}) (Newman *et al.*, 2013). A value of 10 Mbp for the median stall distance was also consistent with the increase in chromosome loss when replication origins were systematically deleted, it was observed that synthetically creating a ~160 kbp origin-less region, by the removal of five efficient origins, increased the chromosome loss rate by ~6 x 10^{-5} (Theis *et al.*, 2010). A simulation of a replicon of the same size and with a

median fork stall distance of 10 Mbp failed to complete replication at a rate of ~6.1 x 10^{-5} (Newman *et al.*, 2013).

Analysis of replicon sizes in *Arabidopsis*, *Drosophila* and human cells by Mamun *et al.* 2016 further test the likely value of N_s. Mamun *et al.* point out that the largest replicon sizes are likely to be bounded by the median stall distance, consistent with an N_s value of 10 Mbp in human cells. Parameter scanning of N_s reveals that a value of 5 Mbp or less saturates the frequency of a cells with at least one DFS (figure 8A of Mamun *et al.* 2016). There are diminishing returns from increasing N_s beyond 10 Mbp, even up to 30 Mbp a DFS would still occur in 20% of S phases. Interestingly, the probability of one, two or three DFSs occurring showed a very pronounced maximum for N_s around 10 Mbp (Al Mamun *et al.*, 2016), consistent with the idea that metazoans have acquired mechanisms for dealing with only a small number of DFSs. Given the similarity of these estimates derived by independent methods, and the conservation of the DNA replication biochemistry from budding yeast to human cells, N_s shall be considered as ~10 Mbp in human cells.

In both models there exists a 'hierarchy of contribution', where certain parameters have a greater effect on the predicted probability of DFSs than others. As indicated in figure 2B, N_g is the greatest contributor to the frequency of DFSs and presents a significant barrier to increasing genome sizes from megabases (*S. cerevisiae*) to gigabases (human cells), even if the frequency of origins is maintained. Origin number is the next greatest contributor (equivalent to decreasing N_l in the earlier DFS model and increasing K in the central equation), but has limited ability to compensate for increasing genome sizes. For example, using the earlier DFS model, fully compensating for the increase in size from the diploid *S. cerevisiae* genome (~22 Mb) to the diploid human genome (~5.6 Gbp) by addition of origins alone would require ~250 fold more licensed origins than already exist in human cells, which would mean the average replicon size would become ~100-200 bp, a similar frequency to histones. Finally, origin distribution is the smallest contributor. The *S. cerevisiae* origin mapping dataset has a coefficient of variation of ~0.7 and so $(1 + R^2) = 1.49$, meaning that

the significant investment in perfectly spacing origins would only reduce DFSs by approximately a third (Newman *et al.*, 2013).

Mamun *et al.* 2016 hypothesise that as genome sizes increase from megabases to gigabases the effect of optimising origin distribution is significantly diminished. So, while yeast origin distributions with a coefficient of variation that is biased towards order (<1), human cells have a coefficient of variation that is more disordered than randomly positioned origins (>1), as there is no evolutionary benefit realised by optimising origin distribution. The diploid human genome (~5.6 Gbp) is ~250 fold larger than the diploid *S. cerevisiae* genome (~22 Mb) and yet has a similar frequency of origins (one origin for every ~31 kb, to *S. cerevisiae's* ~14 kb) which are more disordered (*R* = ~1.5 to *S. cerevisiae*'s ~0.7) (Al Mamun *et al.*, 2016, Newman *et al.*, 2013). This seemingly counterintuitive disparity implies that higher eukaryotes have acquired cellular mechanisms that are able to resolve unreplicated segments of DNA, as even a single DFS in an S phase can potentially cause mis-segregation during the following anaphase. Such mechanisms would relieve the evolutionary pressure from compensating for gigabase genomes by licensing origins at a similar frequency to histones, or optimising origin distribution, and explain why there is no increase in origin frequency or order in human cells when compared to yeast.

The predicted frequency of DFSs in human cells is calculated by the central equation (figure 2B) using the origin-mapping data for HeLa and IMR-90 cells from Picard *et al.* 2014. Centromeric regions, which correspond to the largest replicon in each chromosome in both datasets, are disregarded due to the technical problems associated with sequencing those regions (Al Mamun *et al.*, 2016). IMR-90 cells are a human female primary fibroblast that are stable and untransformed, with a diploid karyotype of 2n = 46 (Nichols *et al.*, 1977), and are included in this analysis to illustrate that the theory applies equally to normal and cancer-derived cells. According to the central equation, the predicted frequency of DFSs in a HeLa cell S phase is ~1.5 (figure 2C, '100% replication origins'). This is more than three orders of magnitude greater than the frequency

predicted for *S. cerevisiae* cells (~1 in 1000) and if left unresolved would mean that human cells risk mis-segregation in most cell cycles. The predicted frequency of DFSs is similar between IMR-90 and HeLa cell line origin-mapping datasets (figure 2C). HeLa cells are highly transformed and have a super-triploid karyotype, but despite these chromosomal rearrangements have maintained a similar genome size, origin number and order compared to the primary cell line IMR-90 (table S2 of Mamun *et al.* 2016). DFSs are not expected to be unique to cancer cells, and they are equally likely to occur in unperturbed, normal cells. In this chapter, *in vivo* experiments are carried out with HeLa cells in concordance with the high-quality HeLa origin mapping data published in Picard *et al.* 2014, but chapter 4 begins with a brief analysis in IMR-90 cells that indicates that the theory is indeed appropriate for human cancer and normal human cells alike.

The central equation is used to predict the frequency of DFSs in response to randomly adding or removing origins from the HeLa and IMR-90 origin-mapping datasets (figure 2C). The number of origins in wild type cells are defined as '100%' (~180,000 from both data sets). Intuitively, randomly removing origins causes the predicted frequency of DFSs to increase, as this will cause the replicons to increase in average length and so the remaining replication forks will be required to replicate more base pairs of DNA before irreversibly stalling (figure 2C, left-hand side). However, randomly removing a significant number of origins has only a modest effect on the predicted frequency of DFSs. For example, randomly removing 40% of origins (from ~180,000 to ~108,000) will only increase the predicted frequency of DFSs from ~1.5 to ~2. This is likely due to the relationship between the average replicon size of ~31 kbp and the median fork stall rate of ~10 Mbp (N_s), hypothetically one replication fork could replicate a median of ~330 replicons-worth of DNA. This redundancy of licensed origins allows cells to tolerate replication stress caused by blockers of fork progression, whether caused by proteins tightly bound to the DNA, DNA damage or collisions with transcriptional machinery, as well as by activating origins that would otherwise remain dormant (Blow and Ge, 2009, Ge *et al.*, 2007, Ibarra *et al.*, 2008, Woodward *et al.*, 2006). The excess of origins, when considering the estimate fork stall rate of ~10 Mb, also protects the genome

against under-licensing, as seen in figure 2C. There is a marked increase in the predicted frequency of DFSs, from ~2 to ~3 when the proportion of randomly removed origins transitions from 40% to 60%, possibly as the system reaches a state of minimal licensing, as the ~30% of origins that are normally activated must now be sufficient for complete replication of the genome.

Conversely, as origins are added at random positions in the HeLa and IMR-90 origin-mapping datasets, the frequency of predicted DFSs decreases (figure 2C, right-hand side). Again, the reduction is modest and illustrates how compensating for gigabase genomes by extraordinary licensing is inefficient. Licensing an additional 50% origins (from ~180,000 to ~270,000), which reduces replicon length from ~31 kbp to ~21 kb, only reduces the predicted frequency of DFSs by approximately a half, which is still orders of magnitude greater than in a typical megabase genome such as *S. cerevisiae*.

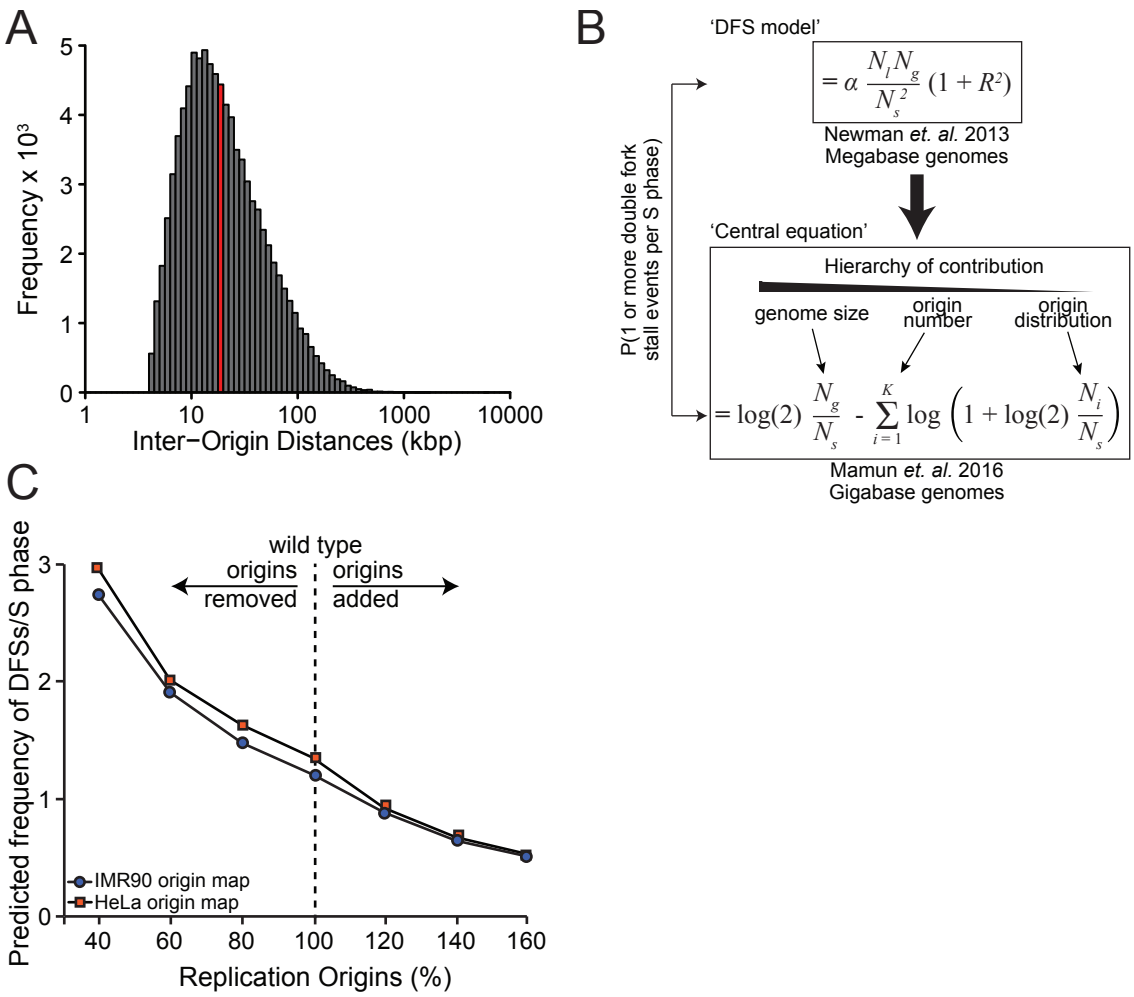


Figure 2: Predicting the frequency of double fork stalls in unperturbed human cells. A) Frequency distribution of ~90,000 replicon lengths (kbp) in HeLa cells, derived from the origin mapping data from Picard *et. al.* 2014, analysed by Dr Mohammed Al Mamun and Dr Luca Albergante. The average replicon length of ~31 kb is indicated by the red bar. **B)** The double fork stall model updated for use with HeLa origin mapping data. N_g = genome size (bp), N_l = average replicon length (bp), N_s = median number of bases replicated before an irreversible replication fork stall (bp), K = number of replication origins in the dataset, N_i = size of particular replicon (bp), *R* = coefficient of variation. The hierarchy of contribution shows which parameters have the greatest effect on the predicted frequency of double fork stalls. Both models were developed by Prof. Timothy Newman with Dr Mohammed Al Mamun and Dr Luca Albergante. **C)** A graph of the effect of randomly adding or removing origins to the genomes of two human cell lines, one cancerous and highly transformed, HeLa, and the other a primary cell line, IMR-90, on the predicted frequency of double fork stalls. Orange squares represent the random addition and removal of replication origins from the HeLa origin mapping dataset (Picard *et. al.* 2014). Blue circles represent the random addition and removal of replication origins from the IMR-90 origin mapping dataset (Picard *et. al.* 2014). Simulation carried out by Dr Mohammed Al Mamun and Dr Luca Albergante.

3.2) G1-specific 53BP1 nuclear bodies: a potential marker for DFSs

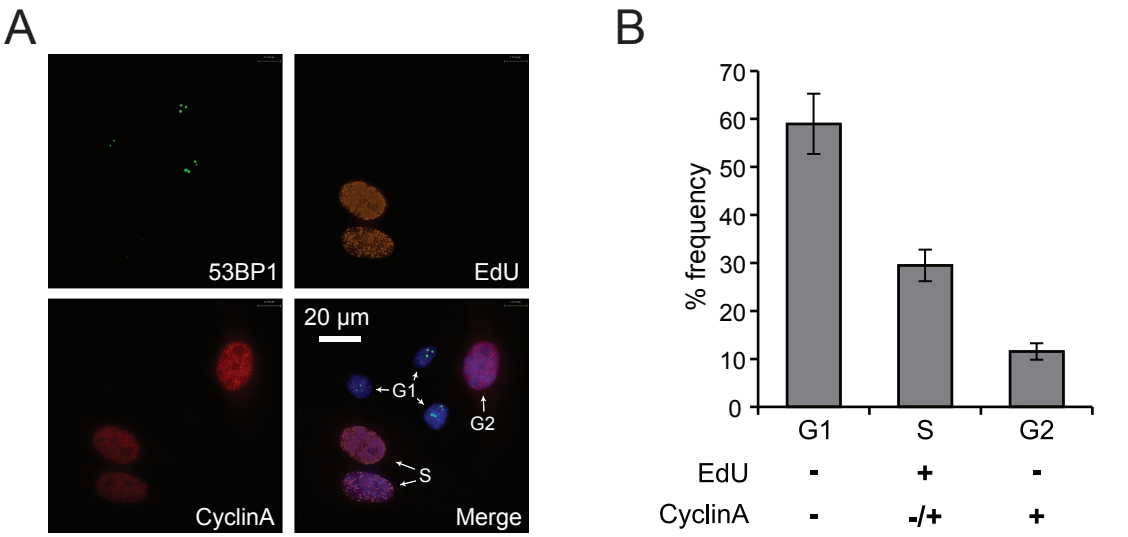
Now armed with the central equation and the well-defined hypotheses on the frequency of DFSs, and how that frequency changes with origin number, a biological marker for DFSs was required to validate the model. Ideally, DFSs would have a unique mark so that they could be stained *in situ* and imaged using fluorescence microscopy, so that their frequency could be quantified per cell. Indeed, there are a number of known proteins that are recruited to stalled replication forks (FANCD2, BLM, MUS81, etc.) but it was not clear whether these markers could be used to differentiate between reversible and irreversible stalls, or between single fork stalls and DFSs. In theory, the segment of unreplicated DNA within a DFS can range from less than the smallest replicon (<1 kb) up to the largest replicon (~6 Mb), although the probability distribution of DFSs is skewed towards larger replicons. So, within a DFS, the physical distance between one fork stall to the other could range from small enough to be unresolvable by wide-field microscopy, to being far apart enough to be undifferentiable from single fork stalls. Alternatively, a DFS could be distinguished by its segment of unreplicated DNA. However, identifying what could be a small segment of DNA amongst the entire genome presents a signal-to-noise ratio problem. For example, detecting unreplicated DNA by incorporation of fluorescently labelled nucleosides (such as bromodeoxyuridine) would act as a negative stain, however, this absence of signal would be swamped by the abundance of nucleotide incorporation that takes place during S phase.

It has long been speculated that unreplicated DNA that persists through to the following anaphase will topologically link sister-chromatids and either cause mis-segregation or require excision or unwinding of the unreplicated DNA (Chan *et al.*, 2007, Clarke *et al.*, 1993). With little to biochemically distinguish a DFS from similar phenomena, quantifying the consequences of transmitting unreplicated DNA through mitosis may act as a good proxy.

The rationale for investigating 53BP1 nuclear bodies as a potential marker for the frequency of DFSs was inspired by the proposition discussed in Harrigan *et al.* 2011 and Lukas *et al.* 2011 that

unreplicated segments of DNA are transmitted through mitosis and form UFBs. These UFBs are either broken or unwound by BLM helicase, and the resultant DNA damage is inherited by each daughter cell in the following G1.

As mentioned in the introduction, under physiological conditions 53BP1 nuclear bodies are largely confined to G1 phase in human cells, as indicated by mutually exclusive immunofluorescence signal to CyclinA (Harrigan *et al.*, 2011, Lukas *et al.*, 2011). However, they have been observed at a lower frequency in early S phase and G2 (Lukas *et al.*, 2011). Ensuring that 53BP1 nuclear bodies are quantified in G1 will limit interpretation to damaged DNA that has been inherited from the previous cell cycle, as opposed to breaks that are generated by internal cellular processes such as DNA replication. As in Harrigan *et al.* 2011 and Lukas *et al.* 2011, CyclinA is used to negatively identify G1 cells, but as its expression progressively builds from late G1 the distinction between G1 and early S phase cells by CyclinA immunofluorescence is not sharp. Short pulse-incorporation of the modified thymidine EdU (5-ethynyl-2'-deoxyuridine) in live asynchronous cells, that can then be conjugated with a fluorophore, allows the identification of S phase cells with a sharp cut-off between G1 and S. This also means that the three stages of interphase can be easily distinguished under the microscope, without having to synchronise the cells, and will from now on be referred to as the '2-stain' system. Mitotic cells are disregarded as 53BP1 nuclear bodies are rare in mitosis since chromatin condensation in mitosis represses ATM and DNA-PK activity and typical DSB repair factors such as BRCA1 and 53BP1 are not recruited to DSBs (Giunta *et al.*, 2010, Nelson *et al.*, 2009). Figure 3A is a representative image of asynchronous HeLa cells stained for 53BP1, CyclinA and EdU. As expected, 53BP1 nuclear bodies (green) can be seen in cells negative for both CyclinA and EdU, defining them as G1. To confirm the accuracy of the 2-stain system, it was used to quantify the frequency of each HeLa cell cycle stage (figure 3B). As indicated in figure 3A ('Merge' panel, white arrows), cells negative for both EdU and CyclinA signal are defined as G1, cells positive for EdU are defined as S, and cells positive for CyclinA only are defined as G2. ~60% of asynchronous HeLa cells are in G1, ~30% in S and ~10% in G2 (mitotic cells disregarded). The



C

EdU	-	+	-	
CyclinA	-	-/+	+	
	G1	S	G2	Total
2-stain staging (% frequency)	**59.0**	**29.5**	**11.5**	100%
HeLa cell cycle timing (h)	~8	~4	~2	14
HeLa cell cycle timing (% frequency)	**57.1**	**28.6**	**14.3**	100%

Figure 3: 2-stain microscopy system to identify G1 cells in an asynchronous population. A) Representative microscopy image of 53BP1 nuclear bodies in the context of the 2-stain staging system. With 53BP1, asynchronous HeLa cells are also stained with EdU (orange) and CyclinA (red). G1 cells can be identified by the absence of both EdU and CyclinA staining. **B)** The proportion of asynchronous HeLa cells categorised as either G1 (EdU -ve/CyclinA -ve), S (EdU +ve) or G2 (EdU -ve/CyclinA +ve). Error bars represent S.E.M. of three biological replicates. **C)** Comparison between the 2-stain staging system and the approximate known lengths of the HeLa cell cycle stages. Mitosis is disregarded as mitotic cells were not categorised by the 2-stain system. χ^2 test (between observed 2-stain staging, and expected stage occupancy), p = 0.726.

approximate HeLa cell cycle times are G1: 8 hours, S: 4 hours, G2: 2 hours and M: 2 hours for a total doubling time of ~16 hours (as calculated in Materials & Methods). Disregarding mitosis, the occupancy of asynchronous cells in each cell cycle stage can be calculated (figure 3C). For example, G1: ~8 h / (~8 h + ~4 h + ~2 h) x 100% = 57.1% of asynchronous cells are in G1. Comparing the observed cell cycle by the 2-stain system (top row of figure 3C) to the calculated occupancy of cell cycle stages by asynchronous HeLa cells (bottom row of figure 3C) indicates that the 2-stain system accurately identified HeLa cell cycle stages.

As applies here and throughout the work presented in this thesis, 53BP1 nuclear bodies were identified for quantification using intranuclear background-corrected signal intensity with a minimum size guide of 0.1 μm^3 (to distinguish nuclear bodies from punctate foci). The distribution of 53BP1 nuclear bodies throughout interphase was quantified by the 2-stain microscopy system in figure 4A. In G1 HeLa cells, the mean frequency of 53BP1 nuclear bodies is ~1. As previously mentioned, 53BP1 nuclear bodies are indeed present in S phase and more infrequently in G2. The frequency of predicted DFSs in unperturbed HeLa cells is ~1.5 (figure 2C, '100% origins') and is well matched by the frequency of G1-specific 53BP1 nuclear bodies. The frequency distribution of G1-specific 53BP1 nuclear bodies is displayed in figure 4B. ~45% of G1 HeLa cells had no 53BP1 nuclear bodies, so approximately half of all unperturbed G1 HeLa cells contain 53BP1 nuclear bodies, which is a significantly higher frequency than expected from unperturbed CFSs, which ranges from 1-5% of cells with observable breaks (Aula and von Koskull, 1976). The overall frequency of G1-specific 53BP1 nuclear bodies is used to construct a Poisson distribution, a distribution where the mean is equal to the variance, and so only requires one variable, λ. The ideal Poisson distribution is super-imposed as a black line over the observed frequency distribution (figure 4B). The observed frequency distribution is highly similar to the Poisson distribution, implying that the observable events are rare and independent. In the context of replication fork stalling, this is an accurate description of DFSs, as the diploid HeLa genome is estimated to contain ~180,000 origins (Al Mamun *et al.*, 2016, Picard *et al.*, 2014) of which ~30% stochastically activate

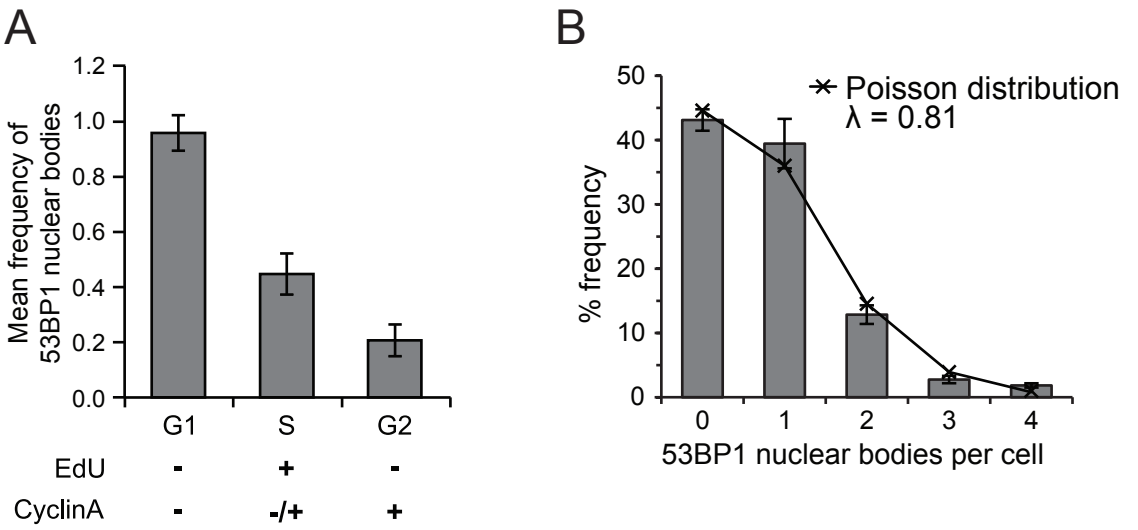


Figure 4: G1-specific 53BP1 nuclear bodies are a potential marker for double fork stalls from the previous cell cycle. A) Mean frequency of 53BP1 nuclear bodies in cell cycle stages defined by the 2-stain staging system. Error bars represent S.E.M. of three biological replicates. **B)** Frequency distribution of G1-specific nuclear bodies in asynchronous HeLa cells. The mean frequency of nuclear bodies was assigned as the λ value to create a Poisson distribution (overlaid black line). Error bars represent S.E.M. of three biological replicates. χ^2 test, p = 0.811.

in a given S phase (Blow and Ge, 2009, Blow *et al.*, 2011, Ge and Blow, 2010, Ge *et al.*, 2007, Ibarra *et al.*, 2008), so there are ~100,000 replication forks of which only 2, 4, 6 or 8 (HeLa cells were not observed to contain five or more 53BP1 nuclear bodies) converging forks stall in approximately half the cells. Additionally, in an unperturbed cell, a spontaneous DFS occurring at one genomic loci should have little-to-no influence on the probability of another DFS occurring elsewhere in the genome. So far, G1-specific 53BP1 nuclear bodies promisingly correlate well with the predicted frequency of DFSs, but further validation is required.

3.3) The temporal distribution of 53BP1 nuclear bodies is constant throughout G1 and declines during S phase

Whether G1-specific nuclear bodies facilitate NHEJ or suppress HR on the products of unreplicated DNA transmitted through mitosis until S phase remains unclear. NHEJ has been reported as a fast process, taking as little as 30 minutes in HCA2 human cancer cells (Mao *et al.*, 2008a). If 53BP1 facilitates NHEJ at 53BP1 nuclear bodies, then it would seem likely that they would appear in early G1 (Harrigan *et al.*, 2011, Lukas *et al.*, 2011) and disappear well before the end of G1, however, live microscopy shows that they linger from at least 80 minutes (Harrigan *et al.*, 2011) to 150 minutes (Lukas *et al.*, 2011). 53BP1 nuclear bodies are predominantly specific to G1, but are present at a low frequency in S phase (Harrigan *et al.*, 2011, Lukas *et al.*, 2011). So, it seems likely that if a cell were to have at least one 53BP1 nuclear body, the body has the potential to appear in early G1 and dissolute in early S phase. To quantify the mean frequency of 53BP1 nuclear bodies across this timespan, HeLa cells were synchronised with the microtubule poison nocodazole, which holds the cells in metaphase (figure 5A). Cells were then released and samples harvested for flow cytometry and microscopy. Figure 5B is a representative example of cells synchronised at mitosis (‘0 h’), entering G1 (‘2 h’) and progressing through G1 until they synchronously enter S phase (’12 h’). 30 minutes before harvesting, sample time-points were incubated in EdU to stage the cells through the cell cycle (figure 6A). At 0 hours from nocodazole release, a majority of the cells have a G2/M (4n) DNA content. 2 hours later a majority of the cells have entered G1, with a 2n total

DNA content. The very first cells start to enter S phase 6 hours after release and continue to do so until the final time-point. The cell cycle stages were defined as presented in the key, and the % frequency of cells were plotted against time after nocodazole release (figure 6B). A majority of cells are in G1 between 2 hours and 8 hours inclusive after release from nocodazole, and so were used for microscopy analysis to quantify 53BP1 nuclear bodies (figure 6C). G1-specific 53BP1 nuclear bodies were selected for by negative EdU and CyclinA staining, to ensure any stray S or G2 cells were not included in the quantification. The mean frequency of G1-specific nuclear bodies remained constant throughout G1 (figure 6C). Although there is a small decline, it is not statistically significant (χ^2 test, $p = 0.924$).

To quantify 53BP1 nuclear bodies through S phase, a double thymidine block was opted against as the synchronisation would not be precise during early S phase. Instead, the 10 and 12 hour time point releases from the nocodazole experiment were imaged and S phase cells were staged by the EdU incorporation morphology. The pattern of fluorescent nucleotide incorporation changes as cells progress through the replication timing programme. Five main morphologies of BrdU incorporation have been identified, and their order and timing discovered by releasing CHO cells from HU synchronisation and tracing the morphology frequency peaks through time, the same patterns were seen in HeLa cells (O'Keefe *et al.*, 1992). In figure 7A are five representative images of each of these stages in HeLa cells incubated with EdU for 30 minutes. The ideal version of the patterns are displayed underneath and adapted from Gillespie and Blow, 2010. It should be noted that each stage is not equal in length, the pattern timing has been estimated in CHOC-400 cells, stage I: <5%, stage II: ~50%, stage III: ~25%, stage IV: ~12%, stage V: ~12% (Dimitrova and Gilbert, 1999). Figure 7B is a representative image of a HeLa cell exhibiting a stage II EdU pattern, and a 53BP1 nuclear body. S phase-specific 53BP1 nuclear bodies were quantified according to the EdU staging pattern, which revealed that 53BP1 nuclear bodies gradually decline during S phase (figure 7C). The very low frequency of nuclear bodies in stage IV and V S phase cells is likely due to how short (and therefore infrequent) those patterns are, which when combined with an infrequent event probably

means that many hundreds of cells would have to be counted for a true reflection. 53BP1 nuclear bodies decline to almost zero during S phase, and the mean frequency of 53BP1 nuclear bodies increases up to ~0.2 per G2 nucleus again afterwards (figure 4A) as NHEJ is favoured over HR again in mid-to-late G2 (Mao *et al.*, 2008b).

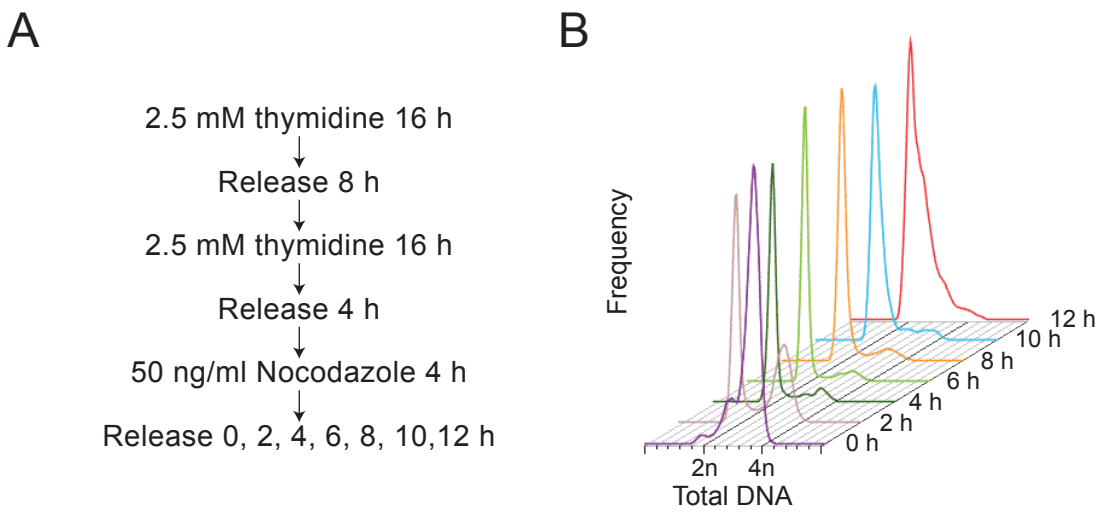


Figure 5: Synchronising HeLa cells for release into G1. A) Schematic of experimental design to synchronously release cells into G1 and track them through to S phase. **B)** Progression of cells released from nocodazole tracked by PI (total DNA) staining.

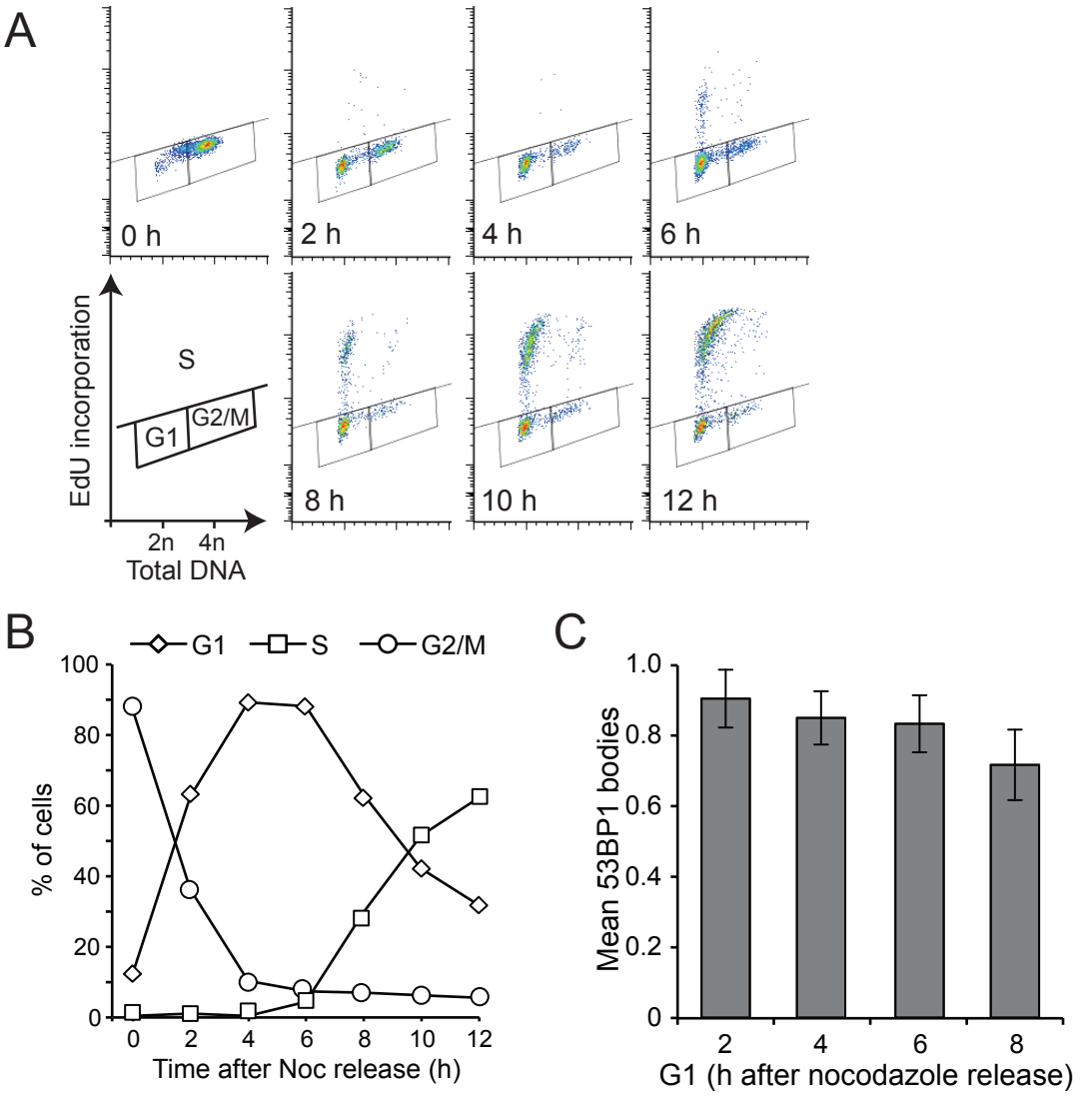


Figure 6: The frequency of 53BP1 nuclear bodies remains constant throughout G1. A) PI intensity (total DNA) plotted against EdU intensity (nucleotide incorporation) at 2 hour intervals after nocodazole release. The key (bottom left) shows the gating used to track progression through the cell cycle after nocodazole release. **B)** % frequency of events in the three gates defined in figure 25A tracked at 2 hour intervals after nocodazole release for a total of 12 hours. **C)** Frequency of G1-specific 53BP1 nuclear bodies quantified at the indicated time points after nocodazole release. Error bars represent S.E.M. of three biological replicates. χ^2 test, p = 0.924.

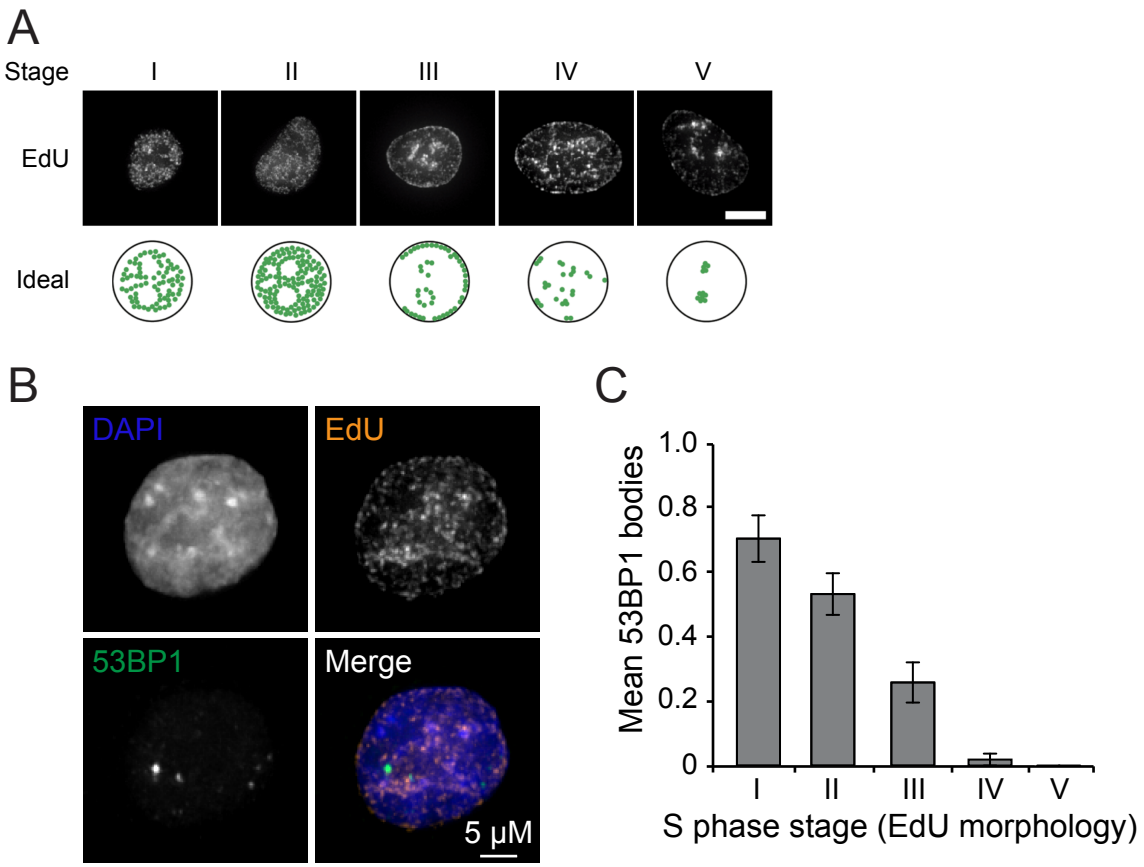


Figure 7: The frequency of 53BP1 nuclear bodies declines during S phase. A) Representative EdU staining of S phase stages as described and defined in O'Keefe *et. al.* 1992. Ideal cartoon displayed underneath from Gillespie & Blow, 2010. **B)** Representative fluorescence image of an S phase HeLa cell (EdU +ve) with a 53BP1 nuclear body. **C)** Quantification of the frequency of 53BP1 nuclear bodies in each S phase stage as indicated by EdU staining pattern. Error bars represent S.E.M. of three biological replicates. χ^2 test, $p = 4.99 \times 10^{-4}$.

3.4) Optimising the depletion of origin licensing by targeting MCM5 and CDT1 with RNAi

The central equation presented in figure 2 not only predicts the frequency of DFSs in unperturbed HeLa cells, but also in cells with randomly added or removed replication origins from the HeLa origin mapping dataset. Intuitively, removing origins from the dataset increases the predicted frequency of DFSs, however, the model goes one step further and predicts the numerical increase with respect to the number of origins randomly removed. If G1-specific 53BP1 nuclear bodies are a good marker for DFSs from the previous cell cycle, then it would be expected that removing origins from HeLa cells *in vivo* would increase the frequency of G1-specific 53BP1 nuclear bodies numerically in line with the models predictions for DFSs.

To reduce the number of origins available for S phase the licensing system was suppressed by depleting MCM5 or CDT1 protein levels. As previously discussed, 53BP1 is an early responder to DNA damage and forms nuclear bodies in response to ɣ-irradiation-induced DSBs (Anderson *et al.*, 2001), a process that is independent of DFSs and replicative stress that subnormal origin activity. This means that the method used to reduce origin usage should be precise and minimise the inhibition of replisome progression, so that any fork stalling detected by 53BP1 nuclear bodies in the following cell cycle should come as a consequence of subnormal origin activation and not directly caused by exogenous agents. This makes targeting the licensing system ideal as the effect will be limited to G1, and should not influence S phase. In contrast, for example, the AID-degron system developed in human cells by Professor Masato Kanemaki and colleagues to rapidly proteolyse MCMs during S phase (Natsume *et al.*, 2016, Natsume *et al.*, 2017) would not be adequate for our purposes as it would be impossible to differentiate 53BP1 nuclear bodies that have accumulated for DFSs from DNA damage caused by fork destabilisation. Targeting a member of the heterohexamer helicase, MCM5, for depletion should have an equivalent effect on licensing as depleting CDT1. The absence of any member of the heterohexamer MCM2-7 eliminates any replication licensing activity for that complex in *Xenopus* cell-free extracts (Prokhorova and Blow, 2000) and elongation in *S. cerevisiae* (Labib *et al.*, 2000). Depletion of MCM5 has previously been carried out in the Blow lab. As can be seen in figure 2D of Ge *et al.* 2007, partial depletion of

MCM5 results in a partial depletion of MCM2, MCM3, MCM6 and MCM7 in the chromatin fraction of U2OS cell lysate (Ge *et al.*, 2007), indicating that fewer MCM2-7 complexes are licensed as a result of limited MCM5 supply. CDT1 promotes the formation of the pre-replicative complex and is required, with CDC6, for the recruitment of MCM2-7 to sites on chromatin marked by the ORC. *S. pombe* CDT1 null-mutants are unable to synthesise DNA (Hofmann and Beach, 1994) and partially depleting CDT1 in *S. cerevisiae* cells blocks initiation of replication and MCM4 (Cdc21) association with chromatin (Nishitani *et al.*, 2000). Depleting either MCM5 or CDT1 result in the same consequence of reduced MCM2-7 on chromatin, and so should both increase the frequency of DFSs and potentially the G1-specific 53BP1 nuclear bodies in the following cell cycle. Additionally, their functional distinction will control for non-specific effects brought about by their depletion. Here, the depletion of MCM5 and CDT1 shall be facilitated by RNAi by transfection of siRNAs that are designed to duplex specifically with their target transcripts, thereby bringing about mRNA degradation and limiting their translation. RNAi can be used to produce a variety of depletions that can then be used to emulate the scatter graph of figure 2C (optimal or hemi-depletion alone would be unsatisfactory as they would only be one data point).

As mentioned above, MCM5 depletion has previously been carried out in the Blow lab and was also facilitated by RNAi. The siRNA sequence for MCM5 used in experiments presented in this thesis is identical to the sequence primarily used in Ge *et al.* 2007 (MCM5-2i). A second sequence was analysed in parallel to MCM5-2i in Ge *et al.* 2007 (MCM5-1i) that achieved a similar depletion of MCM5 protein (~50%) and a similar effect on total nucleotide incorporation and proliferation. Additionally, cells transfected with either siRNA exhibited similar hypersensitivity to HU, but differed slightly on their effect on intra-cluster origin spacing (Ge *et al.*, 2007). The CDT1 siRNA used in experiments presented in this thesis is a SMARTpool (Dharmacon); an equal mixture of four different siRNAs targeted against CDT1 transcripts.

It is necessary to produce a wide variety of origin licensing reductions to emulate the theoretical predictions represented in the scatter graph of figure 2C. In Ge *et al.* 2007, they achieved a total depletion of ~25% MCM5 with 2 nM siRNA transfected, which resulted with a ~50% reduction of MCM2-7 components in the chromatin-bound fraction of U2OS cell lysate (Ge *et al.*, 2007). As can be seen in figure 8A, a similar depletion of ~30% is achieved with 5 nM MCM5 siRNA transfected for 24 hours. However, returns diminish as the concentration of MCM5 siRNA transfected is increased up 20 nM (figure 8A). It seems clear that the concentration of siRNA molecules transfected is not limiting, and it is likely that MCM5 proteins have a long half-life. It is reported that HeLa cell MCM3 and MCM4 have a half-life of ~20 hours (Ishimi *et al.*, 2003, Musahl *et al.*, 1998), which at approximately one HeLa cell cycle implies very little degradation of the protein has occurred and the pool of protein has just been split in two upon cell division. Alternatively, ~50% depletion of an MCM might represent an impenetrable ceiling beyond which transfected cells die, meaning that the cells left for the analysis have been transfected inefficiently. However, this is extremely unlikely: a) a depletion of ~50% of MCM5 in U2OS cells (that also lack an intact licensing checkpoint) has very little effect on S phase dynamics and proliferation rate (Ge *et al.*, 2007), b) ~30% of licensed origins are used in a given S phase (Blow and Ge, 2009, Blow *et al.*, 2011, Ge and Blow, 2010, Ge *et al.*, 2007, Ibarra *et al.*, 2008) and these cells are otherwise unperturbed, c) the theoretical model presented in figure 2C indicates that a significant reduction in licensed origins only has a modest effect on replicative stress. It could not be explained by activation of a licensing checkpoint, as expression of E6 protein from the HPV-18 genome integrated into the HeLa genome means that they lack p53 activity. This means that the potent CDK inhibitor p21 remains inactive and allows constitutive CDK4/6 and CDK2 phosphorylation of Rb (retinoblastoma protein) which means that the E2F transcription factor facilitates expression of S phase cyclins irrespective of licensing status (Blow and Gillespie, 2008, Feng *et al.*, 2003, Montagnoli *et al.*, 2004, Nevis *et al.*, 2009, Shreeram *et al.*, 2002, Teer *et al.*, 2006). In case lipofectamine was the limiting factor, increasing volumes of lipofectamine were mixed with MCM5

siRNAs prior to transfection (figure 8B). Again, a depletion of ~50% was the maximum depletion of MCM5 achievable in 24 hours (figure 8B).

Given the reported slow turnover of MCMs in HeLa cells (Ishimi *et al.*, 2003, Musahl *et al.*, 1998), and that ~50% depletion is seemingly the limit over a 24 hour period, the possibility that greater depletions could be achieved by waiting for cells to divide and share their non-growing pools of MCM5 protein was tested by the development of a transfection regime that covers ~4 HeLa cell cycles (figure 9A). HeLa cells were transfected with either control (non-targeted/scramble), MCM5 or CDT1 siRNA twice over a period of 64 hours (twice, to maintain that the pool of transfected siRNAs within the cells). This regime also proved convenient for creating a variety of depletions, as samples taken every cell cycle (figure 9A) should have progressively less MCM5 protein. MCM5 abundance in whole HeLa cell lysate is indeed progressively depleted after each cell cycle time-point (figure 9B, top). Additionally, depletions beyond ~50% were made possible, as after 64 hours of MCM5 RNAi there is an ~80% depletion of total cell MCM5 (figure 9B, top, right hand lane). However, the objective was to reduce MCM2-7 recruitment to DNA so it is more relevant to quantify depletions in a chromatin-bound fraction of HeLa cell lysate. Similar to the whole cell lysate, depletion of DNA-bound MCM5 is progressive with transfection time (figure 6B, bottom), and the maximal depletion is ~70% after 64 hours of MCM5 RNAi. A similar effect is achieved with CDT1 RNAi (see figure S2A of Moreno *et al.* 2016).

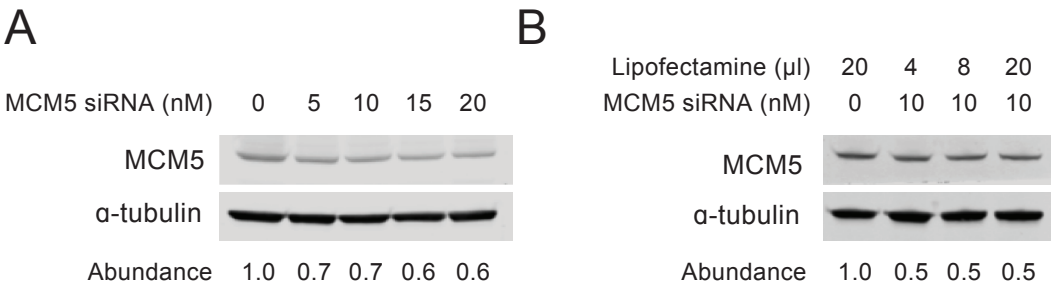


Figure 8: Optimising depletion of replication origins by MCM5 RNAi. A) Immunoblot of MCM5 depletion with increasing concentrations of transfected MCM5 siRNA for 24 hours. Band intensities are normalised against the loading control, α-tubulin. Abundance is band intensity relative to a sample transfected with control siRNA. **B)** Immunoblot of MCM5 depletion with increasing volumes of lipofectamine used on HeLa cells transfected in 100 mm diameter plates. Band intensities are normalised against the loading control, α-tubulin. Abundance is band intensity relative to a sample transfected with control siRNA.

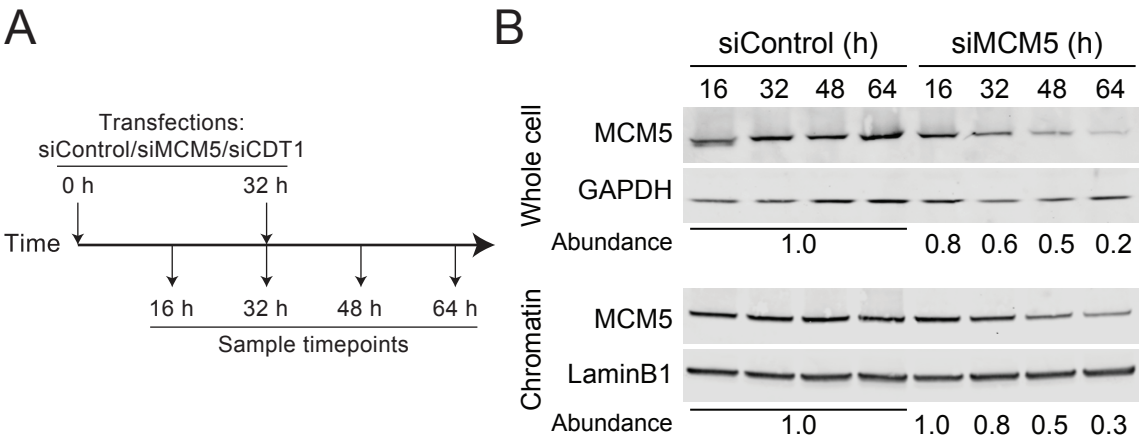


Figure 9: Transfection regime to produce a variety of depletions of MCMs on chromatin. A) Schematic of the experimental design used to create a variety of depletions of MCMs on chromatin. HeLa cells were transfected twice consecutively for 32 hours each, and samples taken every 16 hours until the final sample at 64 hours after the first transfection. **B)** Immunoblots of MCM5 depletion at regular time points for up to 64 hours of RNAi, as described by figure 9A. Band intensities are normalised against the loading controls; GAPDH for whole cell and LaminB1 for the chromatin bound fraction. Abundance is band intensity relative to each equivalent time-point transfected with control siRNA.

3.5) HeLa cells still enter S phase after sever depletion of licensing.

Now that a maximal ~70% depletion of MCMs on chromatin has been achieved in HeLa cells they are somewhat nearing a state of minimal licensing, given that ~30% of licensed origins are used in a given S phase (Blow and Ge, 2009, Blow *et al.*, 2011, Ge and Blow, 2010, Ge and Blow, 2009, Ibarra *et al.*, 2008). This depletion proved to be a second ceiling that was impossible to exceed with RNAi, causing concern that, beyond this level of depletion, HeLa cells were entering S phase, grossly under-replicating and causing large amounts of DNA damage and eventually initiating apoptosis (Blow and Gillespie, 2008, Feng *et al.*, 2003, Shreeram *et al.*, 2002). If this was the case then the cells remaining would be inefficiently transfected, skewing the apparent depletion to be an underestimate. To rule-out this possibility, a flow cytometry technique, developed by Dr Emma Haagensen, designed to observe MCM licensing dynamics through the cell cycle was used to quantify chromatin-bound MCMs specifically in early-S phase cells (the technique was published for the first time in Moreno *et al.* 2016). The basis of the technique is to combine into one sample: total DNA fluorescence (PI), MCM2 immunofluorescence and a 30 minute pulse of EdU fluorescence, in asynchronous HeLa cells with soluble proteins extracted (leaving behind a largely chromatin-bound fraction). Extraction of soluble proteins is necessary as immunofluorescence of MCM2 proteins that are not members of a licensed MCM2-7 heterohexamer, such as a MCM2 homodimer or MCM2, MCM4, MCM6 and MCM7 heterotetramer (Prokhorova and Blow, 2000), will be in excess in MCM5 and CDT1 depleted cells and thereby obscure the immunofluorescent signal from MCM2 as part of a licensed origin.

In figure 10A (top panel), total DNA content is plotted against chromatin-bound MCM2 immunofluorescence intensity in asynchronous HeLa cells transfected with control siRNA for 32 hours (second cell cycle, see figure 9A) and then with EdU for 30 minutes prior to harvesting. G1/G0 phase cells are aligned above an arbitrary intensity of PI (‘2n’) and G2/M phase cells are thus aligned at approximately twice that intensity (‘4n’), with S phase cells situated in between as they progressively increase their total DNA content. Chromatin-bound MCM2 intensity in G1/G0

cells ranges from unlicensed cells (bottom of column) to sufficient licensing to enter S phase (top of column). It is likely that cells with significant chromatin-bound MCM2 intensity signal intensity are G1 cells, as opposed to G0, as quiescent cells do not license origins (Blow and Hodgson, 2002). G2/M cells have little to no chromatin-bound MCM2, as geminin activity strongly inhibits licensing upon the onset of S phase through until the end of mitosis and MCM2-7 helicases (CMGs) dissociate from the DNA as converging forks meet and terminate, which is why MCM2 signal decreases as cells progress through S phase. Background MCM2 signal intensity is slightly higher in G2/M events compared to background intensity in G1/G0 events as the data is uncompensated and there is minor overlap between the emission spectra of PI and Alexa Fluor 488 (MCM2), as well as PI and Alexa Fluor 647 (EdU). In figure 10A (bottom panel) are the same cells but total DNA fluorescence intensity is plotted against EdU fluorescence intensity. Like the MCM2 plot, the G1/G0 and G2/M, both exhibiting background levels of EdU signal intensity, are aligned at the '2n' and '4n' total DNA signal intensities. Cells in the midst of S phase (the peak of the arch shape) are incorporating EdU at the fastest rate, and cells in early and late S phase incorporate EdU at a slower rate (vertical components of the arch. Additionally, cells that enter or exit S phase during the 30 minute EdU pulse will be quantified with a low intensity of EdU as they will have had less exposure to EdU whilst incorporating nucleotides in S phase.

The effect of depleting MCM5 by RNAi for 32 hours can be seen in figure 10B. In the G1 phase population of cells, the peak of chromatin-bound MCM2 immunofluorescence is decreased compared to control. Additionally, the threshold of licensing seemingly required to enter S phase has been reduced. This is likely because of the absence of a licensing checkpoint in HeLa cells (as mentioned previously, the licensing checkpoint does not function in the absence of p53 activity (Montagnoli *et al.*, 2004, Nevis *et al.*, 2009)). MCM5 depletion has not had a significant effect on peak EdU incorporation (figure 10B, bottom panel) as is the case in Ge *et al.* 2007. This is likely because the depletion of origins, although striking in the top panel of 10B, has not pushed the cells into a state of sub-minimal licensing and the ~30% of origins that are used in a given S phase are able to support a normal rate of nucleotide incorporation in these MCM5 depleted cells.

Interestingly, depletion of CDT1 by RNAi for 32 hours does not phenocopy MCM5 depletion (figure 10C, top panel). Rather than a population of cells with a discrete intensity of chromatin-bound MCM2 transitioning from G1 to S phase (figure 10A and 10B, top panels), cells enter S phase with a variety of MCM2 intensities. Speculatively, this could be explained by the difference in how MCM5 and CDT1 are regulated. One MCM5 molecule is required for the licensing of one MCM2-7 helicase, whereas one CDT1 molecule can license multiple origins. Unlike MCM2-7, CDT1 is degraded during S phase and G2 in every cell cycle. Although the depletion of CDT1 is strong (see figure S2A in Moreno *et al.* 2016), it cannot be very high otherwise cells would fail to license a sufficient quantity of MCM2-7, and thus enter S phase and apoptose (Shreeram *et al.*, 2002). The kinetics of CDT1 licensing activity may exaggerate the difference between the most and least efficiently transfected cells. In the less efficiently transfected cells a small number of CDT1 molecules can facilitate normal levels of licensing, and in the efficiently transfected cells a near-zero quantity of CDT1 molecules can only license a small number of origins. The peak rate of EdU incorporation has been slightly diminished in CDT1-depleted HeLa cells (~75% signal compared to control and MCM5-depleted cells, figure 10C, bottom panel). This is probably because a depletion of licensing has resulted in a sufficiently limiting number of replication forks to affect overall incorporation of EdU.

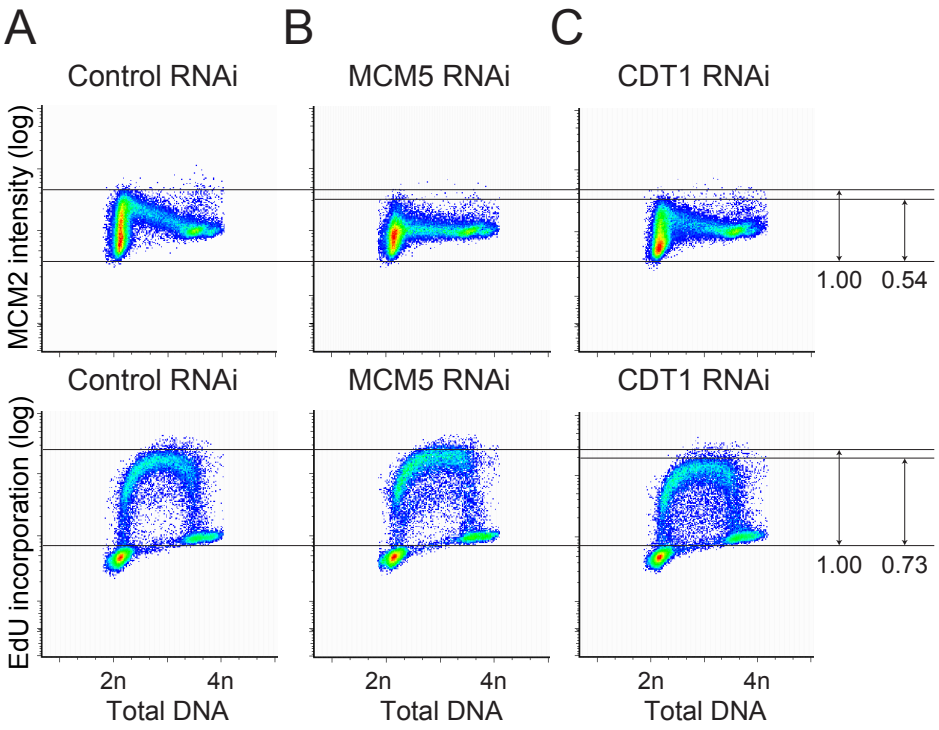


Figure 10: HeLa cells still enter S phase after depletion of MCMs. A) Flow cytometry of the same population of asynchronous HeLa cells transfected with control (scramble) siRNA for 32 hours, twice consecutively. Top: PI (total DNA) plotted against chromatin-bound MCM2 immunofluorescence intensity. Bottom: PI (total DNA) plotted against EdU incorporation intensity of the same cells. **B)** As with figure 10A but cells are transfected with MCM5 siRNA. **C)** As with figure 10A but cells are transfected with

With the flow cytometry system established, it could then be used to quantify the relative depletions of licensed origins in asynchronous early-S phase HeLa cells transfected with either MCM5 or CDT1 siRNA. The EdU flow cytometry profile can be used to separate the asynchronous HeLa cell cycle stages, as illustrated in the left panel of figure 11A. The red population of cells have a 2n content of total DNA (PI) and background levels of EdU fluorescence intensity, defining them as G1/G0 cells. The green population have a 4n content of DNA and background EdU signal, and so are defined is G2/M cells. The blue and orange populations are both S phase cells, defined by EdU signal intensity above the background. The blue population has been distinguished from the rest of S phase cells by gating for cells that are incorporating EdU but have not significantly increased their total DNA content, and defined as early-S phase cells (so the interface between early-S phase cells and late-S phase cells is defined by the right-side extreme of the G1/G0 cells). The same events are replotted with chromatin-bound MCM2 immunofluorescence intensity against total DNA in figure 11A, right panel, maintaining populations defined by the EdU plot. G1/G0 cells (red) exhibit a range of MCM2 intensity, from background levels of licensing at the bottom of the column to sufficiently licensed at the top. Early-S phase cells (blue) are situated at the top of the column, having just entered S phase with a sufficient number of licensed origins. Presented in figure 11B is a schematic of how the flow cytometry system is used to quantify chromatin-bound MCM2 intensity in early-S phase cells. The left panel re-illustrates the gate used to define early-S phase cells, and the events contained within this gate are re-plotted with MCM2 intensity in the middle panel (corresponds to the blue population in figure 11A). The right panel contains the same events as the middle panel, but the MCM2 intensities are plotted as a linear scale. This shows that cells normally enter S phase with a tightly-defined quantity of DNA-bound MCM2.

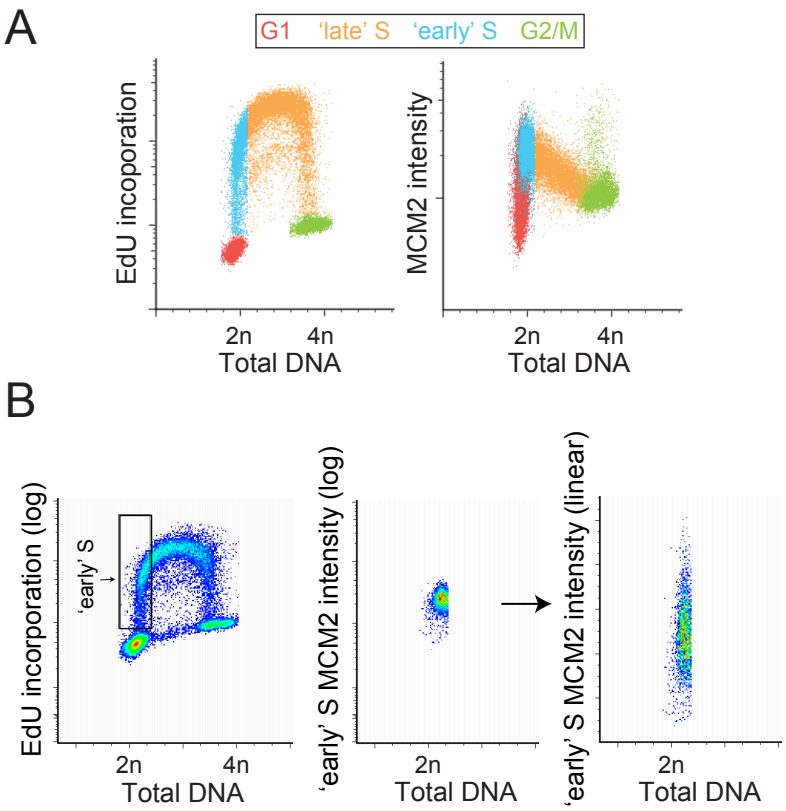


Figure 11: Visualising MCMs on chromatin in 'Early' S phase cells. A) Flow cytometry plots presenting the same events with either EdU intensity or chromatin-bound MCM2 intensity against PI (total DNA). The left plot indicates the cell cycle divided into four categories: G1 (red), 'early' S (blue), 'late' S (orange) and G2/M (green). These same events and their categories are displayed in the right plot. **B)** Control example of selecting 'early' S phase cells for chromatin-bound MCM2 intensity analysis. 'Early' S phase is defined by the gate in the left panel: events that are EdU positive yet not significantly increased in total DNA signal. Middle: 'early' S chromatin-bound MCM2 signal. Right: as with the middle panel but chromatin-bound MCM2 signal intensity is plotted on a linear scale.

The relative immunofluorescence intensity of chromatin-bound MCM2 in early-S phase HeLa cells decreases as they are depleted of MCM5 and CDT1 over a period of 64 hours (figure 12). Asynchronous HeLa cells were transfected with siRNAs twice over a period of 64 hours and samples were harvested every 16 hours, as described in the schematic of figure 6A. The difference between depleting MCM5 and CDT1 is seen here, as in figure 10B and 10C. MCM5-depleted early-S phase cells exhibit a tight cluster of MCM2 signal that steadily decreases to background levels over 64 hours of transfection (figure 12A). In contrast, the effect of depleting CDT1 for 16 hours does not significantly reduce the modal intensity of MCM2, but quickly produces a population of cells with background levels of MCM2 signal. The modal signal does not significantly decrease until 48 hours of transfection with CDT1 siRNA. However, taking the geometric mean shows that using either method to reduce origin licensing has an equivalent effect on early-S phase MCM2 signal intensity.

3.6) Origin number is inversely proportional to the frequency of G1-specific 53BP1 nuclear bodies

With an accurate system to quantify licensing depletion in a population of cells, the cellular response to under-replication can be quantified as the frequency of G1-specific 53BP1 nuclear bodies. After serial transfections with either control, MCM5 or CDT1 siRNA over a period of 64 hours, samples were harvested for flow cytometry and microscopy (figure 9A). Flow cytometry was used to quantify licensing depletion (figure 12), and from the same population of cells immunofluorescence microscopy was used to quantify G1-specific 53BP1 nuclear bodies (figure 3A). In figure 13A and 13B, each data point represents one such population of cells. The cells transfected with control siRNA are defined as having 100% replication origins, and depletions are relative to the corresponding control within a replicate. The mean frequency of G1-specific 53BP1 nuclear bodies in control transfect cells ('100%') is 1.49 ± 0.24 (S.D.), and is slightly higher than untreated cells (figure 4A), likely due to stresses associated with transfection. Depletion of MCM5 is proportional to the frequency of G1-specific 53BP1 nuclear bodies. The maximal depletion

achieved, as measured by relative immunofluorescence intensity of chromatin-bound MCM2 is ~60%, in which cases cells exhibit a mean of ~2.5 G1-specific 53BP1 nuclear bodies, up from ~1.5 in the control cells figure 13A). The effect of CDT1 depletion on the frequency of G1-specific 53BP1 nuclear bodies is subtly greater compared to MCM5 depletion, where the gradient ('M') of the regression is 0.025 compared to 0.019. This is likely due to experimental or biological variation. However, it could also be caused by the heterogeneity of MCM2 loading that is caused by treatment with CDT1 siRNA (figure 12), but could also be a consequence CDT1's alternative function in mitosis. CDT1 was pulled-down by HEC1 (Highly Expressed in Cancer 1), a member of the NDC80 kinetochore-microtubule attachment complex (Varma *et al.*, 2012). As HeLa cells were asynchronous when transfected with CDT1 siRNA, it could be that both under-licensing and mis-segregation events have contributed to the frequency of G1-specific 53BP1 nuclear bodies.

Validation of the theoretical prediction on how origin number influences DFSs (figure 2C) would be significantly strengthened by showing that over-licensing reduces the frequency of G1-specific 53BP1 nuclear bodies. Inducing over-licensing *in vivo* presents a greater technical challenge compared to under-licensing. The licensing system strictly confines licensing to late-M and G1 phase to eliminate the possibility of re-licensing of replicated origins in S phase, which would cause re-replication. Re-replication is disastrous for genomic stability, with defects including mis-segregation/aneuploidy, gene amplification and chromosome translocations (Hook *et al.*, 2007). Overexpression of licensing factors CDT1 and CDC6 induces re-replication and DNA damage in *S. cerevisiae* (Green and Li, 2005), human cancer cells (Vaziri *et al.*, 2003), and makes mice cancer-prone (Munoz *et al.*, 2017). Re-replication is usually only permitted in the absence of p53 activity (Galanos *et al.*, 2016, Vaziri *et al.*, 2003). Professor Vassilis Gorgoulis and colleagues developed a HBEC (Human Bronchial Epithelial Cell) immortalised cell line with a Tet-On doxycycline-inducible promotor for CDC6 overexpression (Petrakis *et al.*, 2016), and gifted the cell line to the Blow lab. Dr Alberto Moreno carried out an experiment where chromatin-bound MCM5 was quantified before and after induction of CDC6 overexpression with doxycycline (figure 14A).

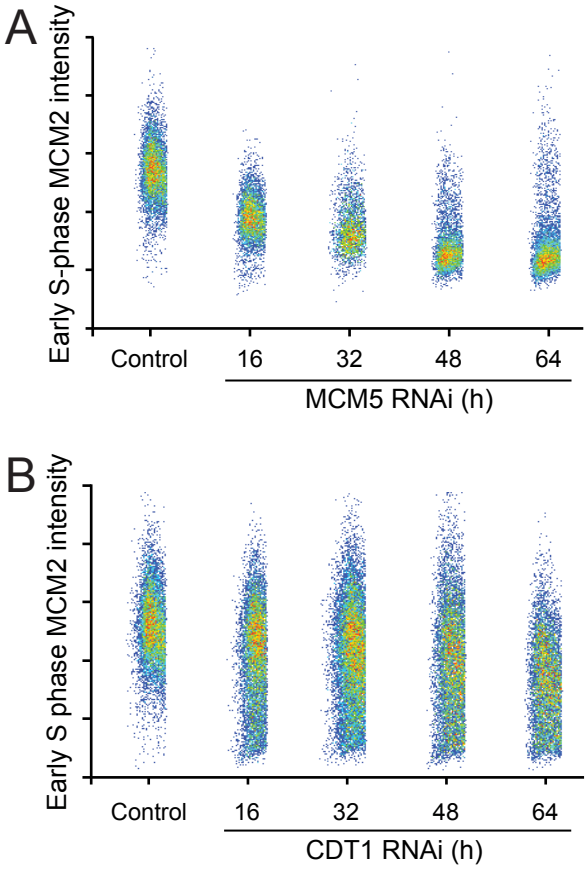


Figure 12: MCMs on chromatin in 'early' S phase cells decrease progressively through the RNAi regime. A) Chromatin-bound MCM2 in early S phase HeLa cells with increasing depletions of MCM5 (see 'early' S population of events in figure 11A, and transfection protocol in figure 9A). **B)** As with figure 12A, but the transcript targeted for depletion is CDT1.

Overexpression of CDC6 has increased the relative amount of chromatin-bound MCM5. Dr Alberto Moreno quantified the frequency of G1-specific 53BP1 nuclear bodies in doxycycline-induced HBEC cells and, as with the experiments presented in figure 13, quantified the relative amount of chromatin-bound MCM5 in the same population of cells.

To verify the theoretical predictions on how origin number affects the frequency of DFSs (figure 2C), calculated licensing reductions (figure 12), and additions (figure 14A), are assumed to be equivalent to the random origin removal and addition to the HeLa origin mapping dataset carried out *in silico*, and the quantified frequency of G1-specific 53BP1 nuclear bodies are treated as a proxy for DFSs. So in figure 14B, the data from figure 13 and 14A is superimposed over figure 2C. On the left-hand side of the '100%' axis are the experiments presented in figure 13, MCM5 RNAi depletions are in red and CDT1 RNAi depletions are in green. The clusters of data points were averaged at every 10% depletion to reduce clutter on the graphs. On the right-hand side is the CDC6 overexpression experiment carried out by Dr Alberto Moreno (light blue). The immortalised HBEC overexpression cell line exhibit ~1 G1-specific 53BP1 nuclear body, and this frequency is reduced to ~0.75 when there's a surplus of ~30% additional MCMs on chromatin. Overall, the three experiments are in excellent concordance with the theoretical predictions.

3.7) 53BP1 preferentially binds to chromatin associated with large replicons

If 53BP1 does accumulate at genomic loci that inherit DNA damage caused by the transmission of unreplicated DNA through mitosis, then it would be expected that these sites would correspond to the largest replicons in the HeLa genome. The probability of a DFS taking place in a given replicon is proportional to its length squared (Newman *et al.*, 2013), which means that they should occur at a disproportionally high frequency at the largest replicons. The increase of chromosome breakage at CFSs caused by aphidicolin treatment is consistent with the idea that a major contributor to their fragility is that they are origin-poor (or where large replicons exist). It was shown by ChIP-qPCR that fragile sites FRA3B and FRA16D are pulled down by 53BP1 and ɣ-H2AX (Lukas *et al.*, 2011).

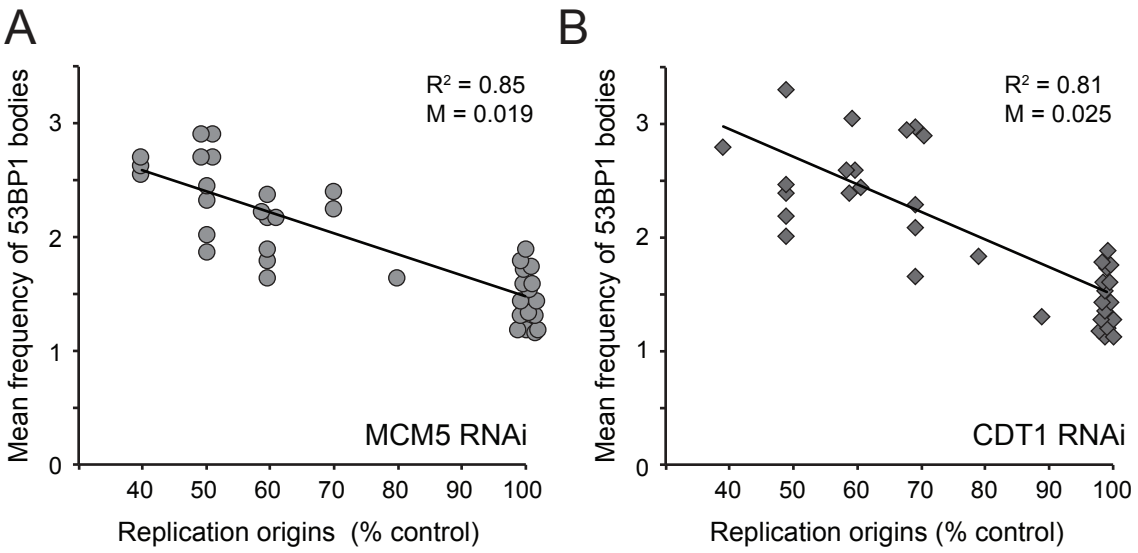


Figure 13: Depletion of MCMs on chromatin increases the frequency of G1-specific 53BP1 nuclear bodies. A) Scatter plot of the depletion of MCM5 as measured by flow cytometry (figure 12) against the frequency of G1-specific 53BP1 nuclear bodies as quantified by fluorescence microscopy (figure 3 and 4). Each data point represents one experiment where both measurements are taken as separate samples of the same population of cells. The line represents the trend using the 'least squares' method (R^2), M is the gradient of the line. **B)** As with figure 13A but depletion of MCMs on chromatin is facilitated by CDT1 RNAi.

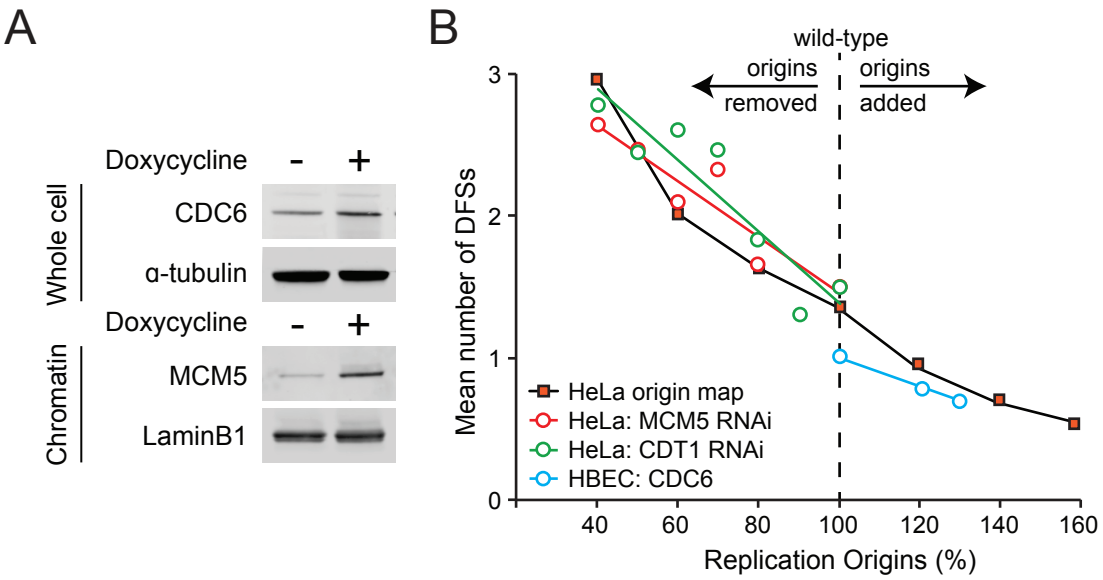


Figure 14: Replicon length is proportional to the frequency of G1-specific 53BP1 nuclear bodies. A) Immunoblots of CDC6 over-expression HBEC cell line before and after induction by addition of 5 µg/ml doxycycline. Induction and blot carried out by Dr Alberto Moreno. **B)** Figure 2C superimposed with the data and trend lines from figures 13A (red) and 13B (green) in addition to G1-specific 53BP1 nuclear bodies quantified in HBEC cells after induction of CDC6 (blue).

This interaction between 53BP1 and ɣ-H2AX with the two CFSs required a treatment with aphidicolin. In unperturbed cells there was no significant enrichment for the fragile sites (Lukas *et al.*, 2011). G1-specific 53BP1 nuclear bodies colocalise with ɣ-H2AX foci, which were shown to bind to ~1 Mbp-worth of chromatin around a DSB (Rogakou *et al.*, 1999). A fraction of the nuclear pool of 53BP1 molecules is bound to chromatin in the absence of DNA damage, although this interaction is transient (Bekker-Jensen *et al.*, 2005). Any regions of chromatin enriched by 53BP1 immunoprecipitation can be sequenced and compared to the distribution of origins from the HeLa origin mapping data, and thereby the distribution of 53BP1 chromatin-binding can be determined relative to small or large replicons.

To confirm this hypothesis with a more holistic approach, where the entire genome can be analysed, Dr Alberto Moreno carried out a ChIP-seq experiment using an anti-53BP1 antibody and IgG (non-specific) control antibody to pull down 53BP1 in unperturbed and asynchronous HeLa cells (Moreno *et al.*, 2016). The precipitated DNA was sequenced and then analysed by Dr Luca Albergante (in Professor Timothy Newman's lab, at the time). The 53BP1/IgG binding ratio is proportional to replicon length (figure 15A). The very largest replicons are few in number (figure 2A), and so the associated 53BP1/IgG binding ratios are relatively noisy (shaded white proportionally to the number of replicons). The correlation between replicon length and 53BP1/IgG is very highly significant ($p < 10^{-15}$), however, the dynamic range of the 53BP1/IgG binding ratio is very small. This could be caused by the high level of stochasticity of pulling down an event that only occurs ~1-2 times per cell cycle and which could potentially occur anywhere in the genome. Alternatively, the high level of background could be caused by the fraction of 53BP1 that is bound to chromatin but not part of a nuclear body. Consistent with this, the genomic coverage of both anti-53BP1 and IgG antibodies are comparable (Moreno *et al.*, 2016). 1 kbp regions of the genome with high 53BP1/IgG binding ratios ($p < 10^{-3}$) were identified, and if one or more these regions fell within a replicon it was defined as a '53BP1+' replicon (Moreno *et al.*, 2016). The average '53BP1+' replicon is three times longer than the average '53BP1'- replicon (figure 15B).

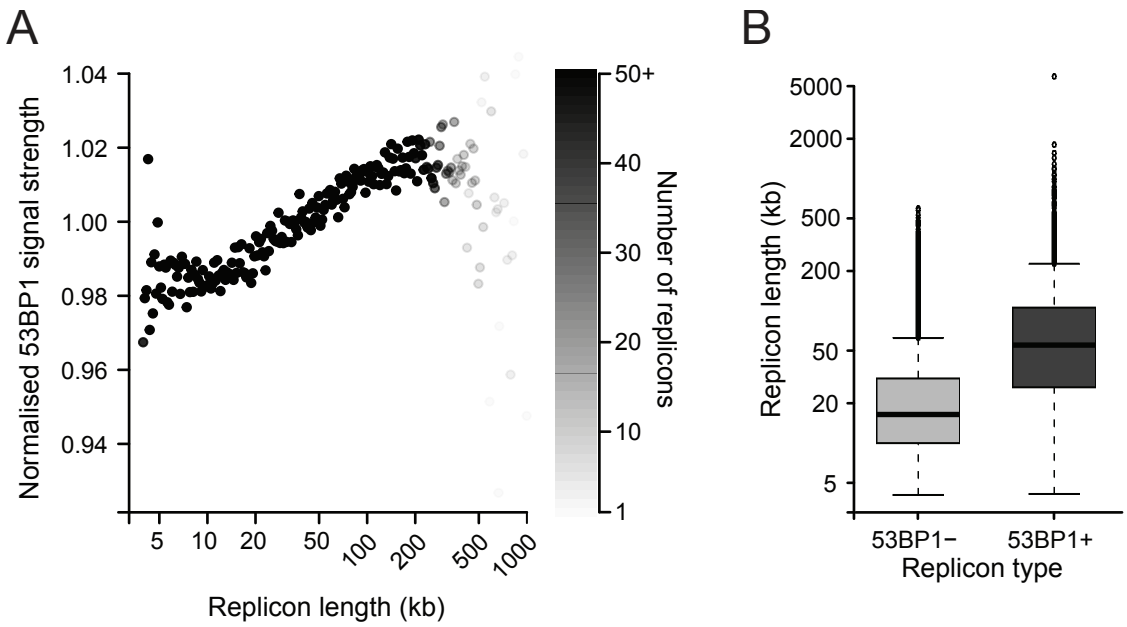


Figure 15: 53BP1 preferentially binds at large replicons. A) Scatter plot of 53BP1/IgG binding ratio per kilobase against replicon size. Replicon sizes were grouped and each data point represents the mean of each group. The shade of each datapoint represents the number of reads in that replicon size group (the largest replicons have very few reads and are thus noisy). Spearman, $\rho = 0.91$, $p < 10^{-15}$. ChIP-seq carried out by Dr Alberto Moreno, and analysis by Dr Luca Albergante. **B)** Mean length of replicons with (53BP1+) or without (53BP1-) significant 53BP1/IgG binning ratios. t-test, $p < 10^{-15}$. ChIP-seq carried out by Dr Alberto Moreno, and analysis by Dr Luca Albergante.

To better our understanding of the nature of the binding of 53BP1 to chromatin, cells were fractionated to determine the proportion of 53BP1 that is chromatin-bound, and microscopy was used to quantify the intensity of 53BP1 nuclear body signal compared to total nuclear signal. U2OS cells were used for these experiments as Professor Jiri Lukas and colleagues have developed a well-characterised and validated U2OS cell line that expresses a GFP-53BP1 fusion protein at near-physiological expression levels; the amount of total protein is equivalent between fusion and endogenous and the fusion protein retained physiological functions (Bekker-Jensen *et al.*, 2005, Jullien *et al.*, 2002). The rationale for using this cell line, as opposed to immunofluorescently labelling endogenous 53BP1, is that the ratio between the fluorescent molecule, GFP, and the ectopic 53PB1 is 1:1, which is not necessarily the case for immunofluorescence, meaning that quantification of GFP fluorescence should be in the linear range. Firstly, U2OS cells were fractionated, producing a chromatin-bound insoluble lysate and a soluble lysate (figure 16A). Equivalent volumes of each lysate were loaded so that abundances of 53BP1 in the chromatin-bound fraction could be quantified relative to the 53BP1 abundance in the soluble fraction. LaminB1 and α-tubulin were also blotted to verify the fractionation. LaminB1 is an intermediate filament protein that is insoluble and remains with the chromatin after fractionation, and α-tubulin is a soluble protein that does not strongly associate with chromatin. The background-corrected signal intensities of the 53BP1 bands (in arbitrary units) are shown beneath the blot. ~80% of all 53BP1 is chromatin bound (figure 16A). Figure 16B is a representative image of the GFP-53BP1 U2OS cell line. When quantifying the fluorescence intensity of 53BP1, normal U2OS cells were imaged under the same conditions to obtain the intranuclear background. Software was used to find the volumes that corresponded to nuclear bodies and the entire nucleus, and background was removed on a per-voxel basis. The relative intensity of GFP-53BP1 in nuclear bodies is ~1% when compared to the total nuclear signal (figure 16C). Taken together with the fractionation experiment (figure 16A), it seems that a majority of 53BP1 (~80%) is bound to chromatin without being part of a nuclear body in unperturbed U2OS cells (figure 16D). This is consistent with the 53BP1 ChIP-seq data that indicated 53BP1 had the potential to bind almost anywhere in the genome and probably explains

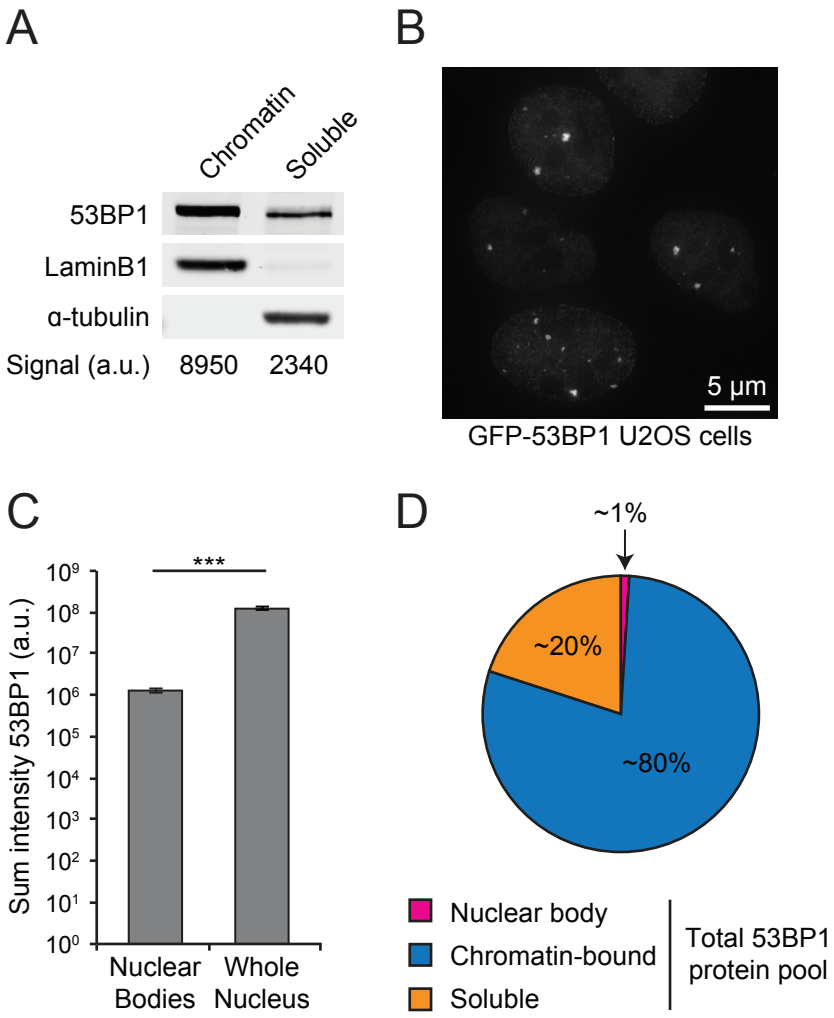


Figure 16: A significant proportion of all 53BP1 proteins are bound to chromatin and not as part of nuclear bodies. A) Immunoblot of 53BP1 in either the chromatin-bound fraction or the soluble fraction, each fraction loaded at the same volume. **B)** Representative image of the GFP-53BP1 U2OS cell line. **C)** Sum voxel intensity of GFP signal in 53BP1 nuclear bodies in comparison the sum voxel intensity of GFP in the whole nucleus. Error bars represent S.E.M. of three biological replicates. t-test, $p = 8.18 \times 10^{-37}$. **D)** Pie chart of the nuclear localisation of 53BP1, summarising the results from figure 16A, B and C.

why the dynamic range in figure 15A is small. Pan-nuclear chromatin-bound 53BP1 with sites of extreme 53BP1 accumulation and activity is reminiscent of the distribution of H2AX variant histone throughout the genome, but upon activation (phosphorylation) accumulates at the local site of damage and forms a nuclear body (Bewersdorf *et al.*, 2006, Natale *et al.*, 2017). It is possible that 53BP1 is similarly poised throughout the genome to function with ɣ-H2AX to repair damaged DNA.

3.8) Dormant origins and 53BP1 function synergistically to protect genomic stability against under-replication.

53BP1 nuclear bodies protect unstable DNA structures inherited from the previous cell cycle (whether a long stretch of ssDNA generated by BLMs-mediated unwinding of a UFB or DSBs created by MUS81-mediated excision of stalled-fork structures at CFSs) by repressing BRCA1-mediated HR. 53BP1-deficient mice are cancer-prone and exhibit stunted growth (Morales *et al.*, 2003, Ward *et al.*, 2003b), and when depleted in BRCA1-/- mice this phenotype is rescued (Bouwman *et al.*, 2010, Bunting *et al.*, 2010), implying that inappropriate HR in G1 cells causes genomic instability. A partial-depletion of 53BP1 in HeLa cells could reduce HR inhibition at 53BP1 nuclear bodies and increase genomic instability and cell death. However, the frequency of DFSs are predicted to be relatively infrequent (~1-2 per S phase (figure 2C)) and there may be redundant pathways that can protect or resolve a small number of DNA damage events. Unperturbed HeLa cells rarely exhibit four or more G1-specific nuclear bodies (figure 4B) and could indicate the low number of DFS events a HeLa cell can tolerate. Al Mamun *et al.* 2016 speculate that if DFSs generate DSBs then the number of illegitimate end rejoinings increases disproportionately as the frequency of DFSs increases, reducing the probability of restoring the correct chromosomal structure. The number of possibilities for NHEJ recombination between DSB ends is approximate to the factorial of the number of DSBs. For example, 2 DSBs create 4 DSB ends that can be rejoined in 3 different ways, however 3 DSBs can be rejoined in 15 different ways and 4 DSBs in 105 different ways. If this is the case then it might be expected that increasing the

maximum frequency of DFSs in a single cell to beyond ~3, in combination with a partial-depletion of 53BP1, will additively increase genomic instability. However, DSB anchoring facilitated by γ-H2AX, MRN, MDC1 and 53BP1 might ensure that DSB ends remain locally associated to preserve the structure of the genome, meaning that illegitimate end rejoining is less likely (Bassing and Alt, 2004).

To test this hypothesis, a clonogenic assay on asynchronous HeLa cells transfected with MCM5 and 53BP1 siRNAs was carried out. Depletions of MCM5 and 53BP1 were confirmed (figure 17A) and the number of colonies that grew in increasing concentrations of HU were quantified (figure 17B). Partial-depletion of 53BP1 had no effect on HeLa sensitivity to HU treatment (figure 17B, green line), indicative of a toleration to replicative stress that is not dependent on high levels of 53BP1 activity. Partial-depletion of MCM5 makes HeLa cells slightly more sensitive to replicative stress (figure 17B, red line, also see figure 5D of Ge *et al.* 2007) as fewer dormant origins are activated upon treatment with HU compared to the activation seen in cells with normal levels of origin licensing (Ge *et al.*, 2007). The cells become most sensitive to HU when both MCM5 and 53BP1 are partially-depleted (figure 17B, blue line). The sensitivity is well beyond the additive effect of individually reducing MCM5 and 53BP1, which implies a synergistic interaction between dormant origins and 53BP1 to promote genomic stability. Partial-depletion of replication origins increases the frequency of DFSs which are inherited as unstable sites or regions of DNA damage that would normally be protected from decay or inappropriate HR. HU treatment also increases the rate of fork stalling, DFSs, and, under normal conditions, the appearance of 53BP1 nuclear bodies. With the reduced abundance of 53BP1 at these loci, BRCA1-mediated HR is likely to take place at a higher frequency. Quantifying the number of colonies in the untreated conditions (by HU), without normalisation, shows that there is no significant difference in cell death after partial-depletion of MCM5, 53BP1, or both (figure 17C). 500 cells were seeded for each plate, which translates to ~300 colonies quantified in the cells transfected with control siRNA, and does not significantly vary after the partial-depletions. This indicates that HU is required for the synergistic interaction to become

evident, likely because the induced increase in DFSs by MCM5 RNAi does not cause sufficient replicative stress to overwhelm the resultant DNA damage response, even if 53BP1 is partially depleted. It is also possible that alternative pathways that facilitate repair of under-replication do so before 53BP1 is required in the following G1 phase. These results are in concordance with an experiment where enrichment of TdT (indicates DSBs) at the origin-poor CFSs, FRA3B and FRA16D, only takes places when cells are both depleted of 53BP1 and treated with aphidicolin (Lukas *et al.*, 2011).

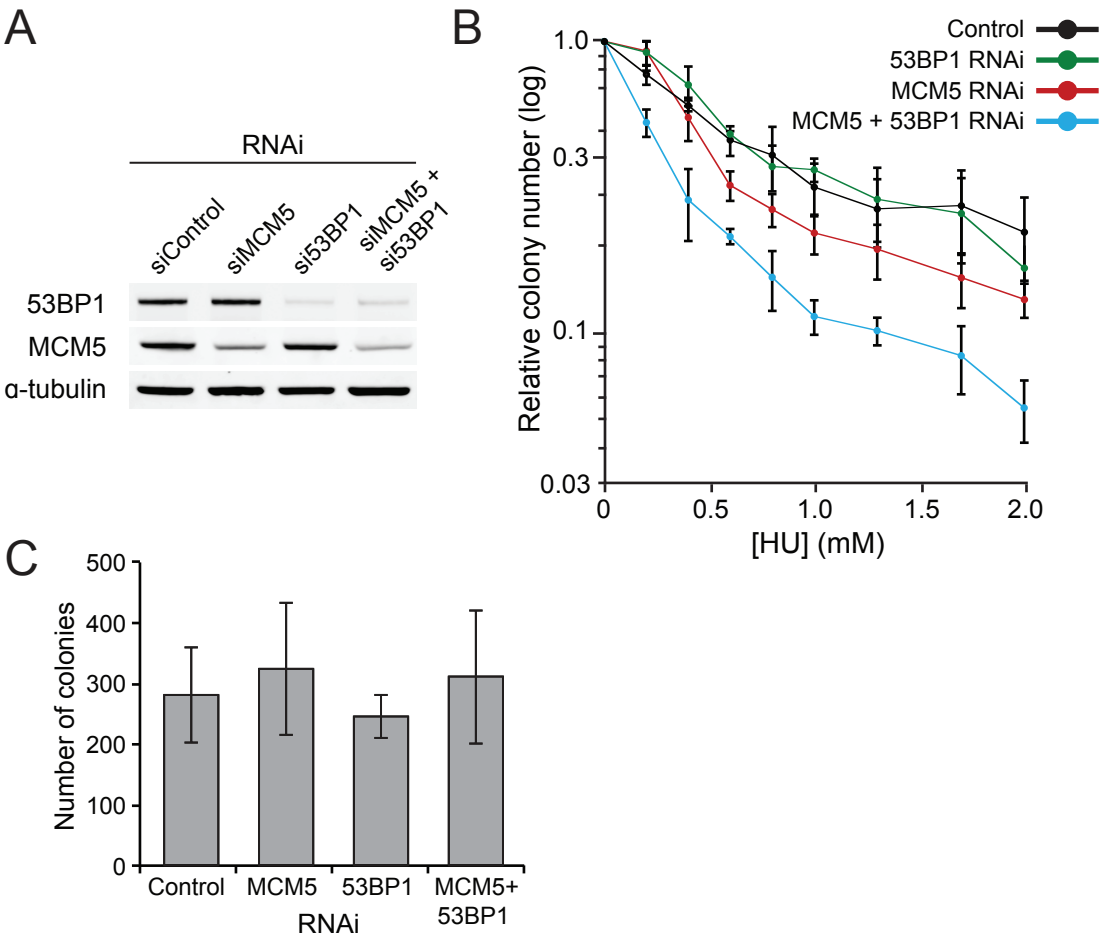


Figure 17: Dormant origins and 53BP1 function synergistically to protect genome stability against under-replication. A) Immunoblot to confirm the depletion of both MCM5 and 53BP1 in HeLa cells used in the experiments presented in figure 17B-C. **B)** Clonogenic assay of HeLa cells depleted of either 53BP1, MCM5 or both. Colonies formed after treatment with increasing concentrations of HU were quantified and normalised to the untreated cells. Error bars represent S.E.M. of three biological replicates. K-S test, P = 1.00, P = 0.338, P = 0.0366, when comparing the 'Control' distribution to the '53BP1', 'MCM5' and 'MCM5 + 53BP1' distributions respectively. **C)** Plating efficiency of cells for the clonogenic assay. Number of colonies of the control cells not treated with HU, pre-normalisation. Error bars represent S.E.M. of three biological replicates.

Chapter 4

Cellular responses to double fork stalls

4.1) G1-specific nuclear bodies in U2OS and IMR-90 cells

The theoretical model predicts the average frequency of DFSs in a given HeLa S phase, and did so based on the genome size, origin mapping data, and an estimation for the median number of bases a replication fork replicates before irreversibly stalling. As described in Chapter 3, hypotheses that the model produced were confirmed by strong correlation with a candidate protein, 53BP1, that was already known to respond to replicative stress from the previous cell cycle (Harrigan *et al.*, 2011, Lukas *et al.*, 2011). 53BP1 also preferentially binds to large replicons and functions with dormant origins to protect genomic stability. None of these features depend on the cells in question to be cancerous. As long as the total genome size and number and distribution of replication origins are largely conserved between different types of human cells, then the frequency of DFSs should also be conserved. This can be seen in the small difference between the predicted frequencies of DFSs between HeLa and IMR-90 cells in figure 2C. However, there is a possibility that the characteristics of G1-specific 53BP1 nuclear bodies, with respect to the DFS model, presented so far in this thesis are unique to HeLa or cancer cells. HeLa cells are notoriously unstable and highly transformed, and exhibit cancer hallmarks such as gene amplification, deletions, chromosomal rearrangements and loss of heterozygosity (Frattini *et al.*, 2015). The HeLa cells used in the experiments presented in this thesis were supplied from the ATCC (known as HeLa-CCL2) and are reported to have a modal chromosome number of 82, but can range from 70 to 164. However, a study reported that this cell line instead has a modal chromosome number of 78, which can range from 76 to 80 (Macville *et al.*, 1999). The instability is largely thought to be caused by aneuploidy which is characteristic of the HeLa cell line (Nicholson and Cimini, 2013). This instability and aneuploidies could cause DNA damage and activate 53BP1 accumulation at nuclear bodies, independent of DFSs.

In order to test this idea, G1-specific 53BP1 nuclear bodies were quantified in U2OS and IMR-90 cells to confirm that the basic characteristics were similar to those seen in HeLa cells. Human U2OS cells (osteosarcoma), like HeLa are highly transformed, exhibiting a modal chromosome number of ~68 (Janssen and Medema, 2013). However, U2OS cells are relatively stable when

compared to HeLa cells, with a chromosome number only slightly higher than when the cell line was first cultivated (ATCC). U2OS cells are null for the CDKN2A gene (CDK inhibitor 2a) which in normal cells has two gene products, $p16^{INK4A}$ and $p14^{ARF}$. $p16^{INK4A}$ is an inhibitor of CDK4 and CDK6 activities, and its loss means that Rb is hyperphosphorylated and unable to repress E2F transcription factors (Serrano *et al.*, 1993), promoting the expression of genes required for DNA replication. $p14^{ARF}$ (alternate reading frame) is expressed from the same locus and inhibits HDM2-mediated (Mdm2 in *S. cerevisiae*) repression of p53. HDM2 is transcriptionally activated by p53 and also represses p53, in a negative feedback loop by ubiquitylation which exports p53 to the cytoplasm for proteolysis (Zhang and Xiong, 2001). $p14^{ARF}$ binds to HDM2 to inhibit its E3 ubiquitin ligase activity (Honda and Yasuda, 1999) and sequesters it to nucleoli (Weber *et al.*, 1999). So, in the absence of these two factors, U2OS cells lack regulated E2F-mediated transcription of S phase cyclins and this may therefore explain its lack of a licensing checkpoint, as with HeLa cells.

Figure 18A is a representative image of U2OS cells stained as for HeLa cells in figure 3A. Dr Alberto Moreno quantified the frequency distribution of G1-specific 53BP1 nuclear bodies in U2OS cells. There are ~1.5 G1-specific 53BP1 nuclear bodies in U2OS cells (figure 18B), compared to ~1 in HeLa cells (figure 3C). Despite the slightly higher frequency of G1-specific nuclear bodies in U2OS cells, their frequency distribution still conforms to a Poisson distribution (figure 18B, black line). A partial depletion of MCM5 (figure 18C) increased the mean frequency of G1-specific 53BP1 nuclear bodies from ~1.5 to ~2.5 (figure 18D). There is no U2OS origin mapping data that these values can be compared to, but when compared to HeLa they are relatively similar, if not slightly higher, indicating that the link between the DFS predictions and 53BP1 nuclear bodies is valid in at least one other cancer cell line.

IMR-90 cells, in contrast to HeLa and U2OS cells, are a primary fibroblast cell line cultivated from a 16-week old female foetus and are untransformed and stable with a normal chromosome number

of 46 (Nichols *et al.*, 1977). Figure 19A is a representative image of IMR-90 cells stained as for HeLa cells (figure 3A) and U2OS cells (figure 18A). The frequency distribution of G1-specific 53BP1 nuclear bodies was quantified (figure 19B). The overall mean is ~0.5 nuclear bodies and still fits well to a Poisson distribution (black line). Partial depletion of MCM5 (figure 19C) increases the mean frequency of G1-specific 53BP1 nuclear bodies to ~0.9 (figure 19D). Out of the three cell lines where G1-specific 53BP1 nuclear bodies have been quantified, IMR-90 has the lowest frequency. Although the sample size is small, it could be that the chromosomal rearrangements and hypertriploidy as seen in HeLa and U2OS cells causes an increase in G1-specific 53BP1 nuclear body frequency.

4.2) Alternative responses to DFSs: MiDAS and FANCD2

With respect to the cell cycle, there is a long period of time between a DFS in S phase and the accumulation of 53BP1 nuclear bodies in the following G1. It seems unlikely that 53BP1 is the first component of the DNA damage repair response that facilitates synthesis or repair at unreplicated segments of DNA. Although 53BP1 nuclear bodies accumulate rapidly and early in G1 (Lukas *et al.*, 2011), a DFS-induced unreplicated segment of DNA must be preserved from the beginning of G2 through to anaphase where it may be excised by MUS81 endonuclease or form a UFB that is unwound by BLM helicase (Chan *et al.*, 2007, Chan *et al.*, 2009, Ying *et al.*, 2013). Note, one of the core assumptions in the development of the original DFS model was that a population of replication forks stall irreversibly, and is built into the model by the median number of bases replicated before stalling ($N_s \approx 10$ Mb) (Newman *et al.*, 2013). Therefore, fork restart that successfully replicates a previously unreplicated segment of DNA and which can be resolved prior to mitotic entry, is disregarded as a response that is specific to DFSs.

Given the relevance of MiDAS to fork stalling and reducing the amount of unreplicated DNA transmitted through anaphase, it is potentially relevant to resolving DFSs. To test this,

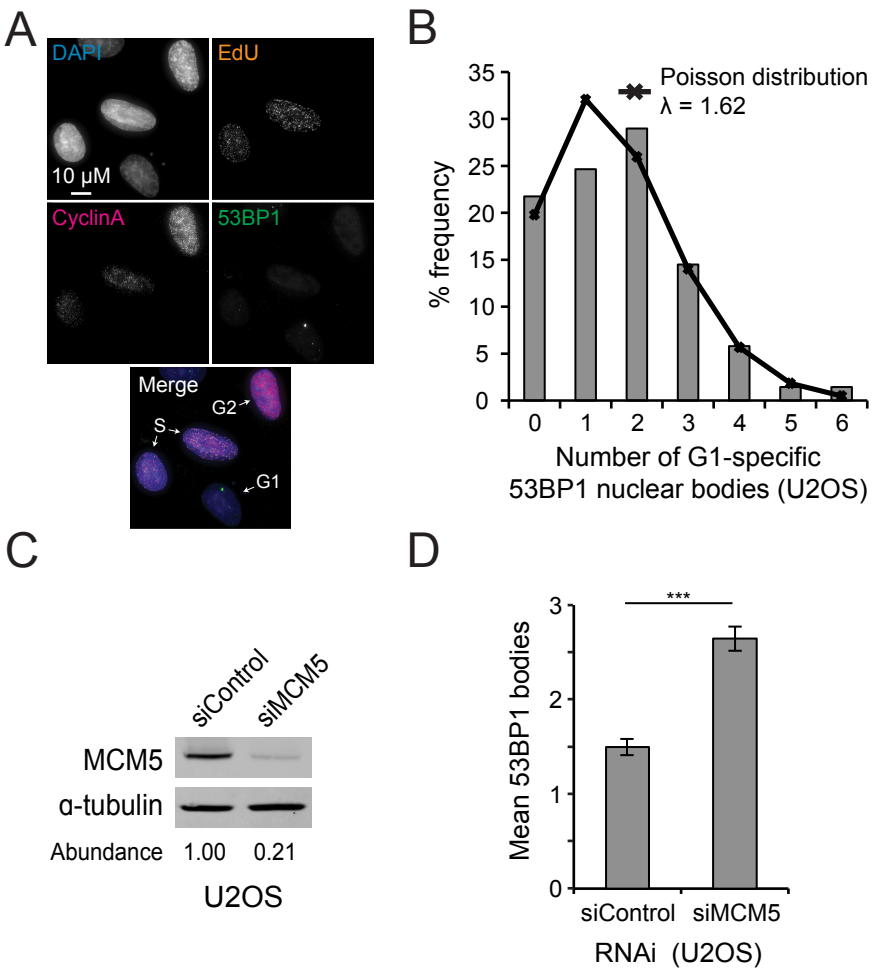


Figure 18: G1-specific 53BP1 nuclear bodies in U2OS cells. A) Representative image of an asynchronous, unperturbed U2OS cell stained for 53BP1 (green), EdU (orange) and CyclinA (pink). **B)** Frequency distribution of G1-specific 53BP1 nuclear bodies in asynchronous, unperturbed U2OS cells. The mean frequency (1.62 nuclear bodies per G1 nucleus) is used to construct a Poisson distribution (black line). χ^2 test, p = 0.651. 53BP1 nuclear bodies imaged and quantified by A.M. **C)** Immunoblot of MCM5 depletion in U2OS cells for the experiment presented in figure 15D. Band intensities are normalised against the loading control, α-tubulin. Abundance is band intensity relative to a sample transfected with control siRNA. **D)** Quantification of the mean frequency of G1-specific 53BP1 nuclear bodies in U2OS cells either transfected with control (scramble) or MCM5 siRNA. Error bars represent S.E.M. of three biological replicates. t-test, p = 1.4×10^{-12}.

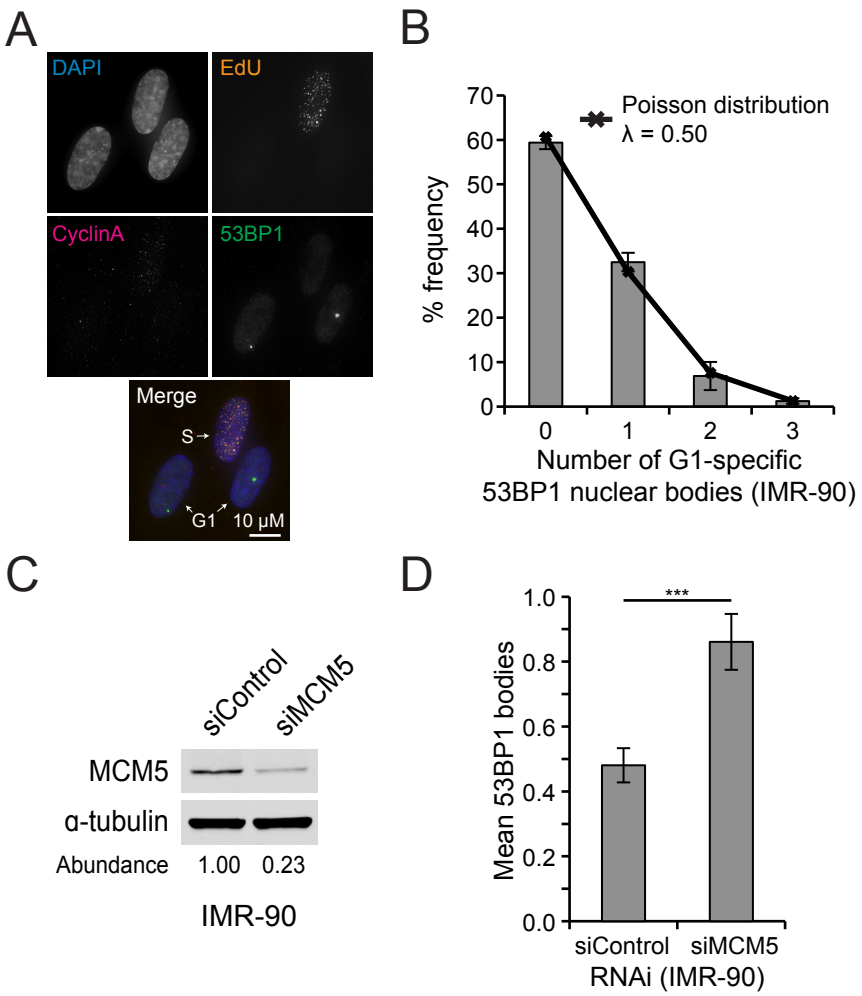


Figure 19: G1-specific 53BP1 nuclear bodies in IMR-90 cells. A-D) As with figure 18A-D, but with IMR-90 cells. χ^2 test for figure 19B, p = 0.969. t-test for figure 19D, p = 1.79×10^{-4}.

MiDAS foci frequency were quantified in response to MCM5 RNAi, for comparison with the numerical predictions generated by the DFS model.

The introduction describes concerns about the use of the CDK inhibitor RO-3306 for assaying MiDAS. Figure 20A outlines a method to detect MiDAS where RO-3306 is not used. Instead, asynchronous U2OS cells are incubated in EdU for 30 minutes before fixation, and prophase and prometaphase cells are sought out using DAPI and p-H3 fluorescence morphology (similar to the method used for figure 1 of Minocherhomji *et al.* 2015). Meraldi *et al.* 2004 and Minocherhomji *et al.* 2015 show that U2OS cells take >30 minutes to reach prometaphase, and in figure 2A of Lukas *et al.* 2011 indicates that it takes ~ 1 hour. The intention of this method is that cells were fixed such that they had entered mitosis more than 30 minutes previously, so that EdU incorporation takes place exclusively in mitosis. Although this assumption could be challenged, the assay uses a more physiologically-true system than the use of RO-3306. Cells were treated with aphidicolin to compare to experiments in Minocherhomji *et al.* 2015. A representative image of an aphidicolin-treated prophase U2OS cell is presented in figure 20B. In addition to fluorescently labelling EdU, FANCD2 is immunofluorescently stained, given the previously mentioned roles of FANCD2 in CFS resolution in mitosis, and the reported colocalisation between MiDAS foci and FANCD2 foci (Minocherhomji *et al.*, 2015). FANCD2 foci are recruited to chromatin when ubiquitylated, and dissociate when deubiquitylated by USP1, after the damage is repaired (Garner and Smogorzewska, 2011, Nijman *et al.*, 2005). This means that in this experiment FANCD2 foci are a comparative snapshot of fork stalls, so mitotic EdU foci that colocalise with FANCD2 partially legitimises them as MiDAS foci. Inset into the 'Merge' panel of figure 20B shows some representative examples of colocalisation between MiDAS foci and FANCD2 foci. Quantification of FANCD2 and MiDAS foci in the experiment described in figure 20 is presented in figure 21. There is a mean of ~5 FANCD2 foci in unperturbed prophase or prometaphase U2OS cells, which increases to ~35-40 after a 24 hour treatment with aphidicolin (figure 21A), and is approximately in line with previously reported quantifications in metaphase nuclei in figure 1F of Chan *et al.* 2009. In unperturbed

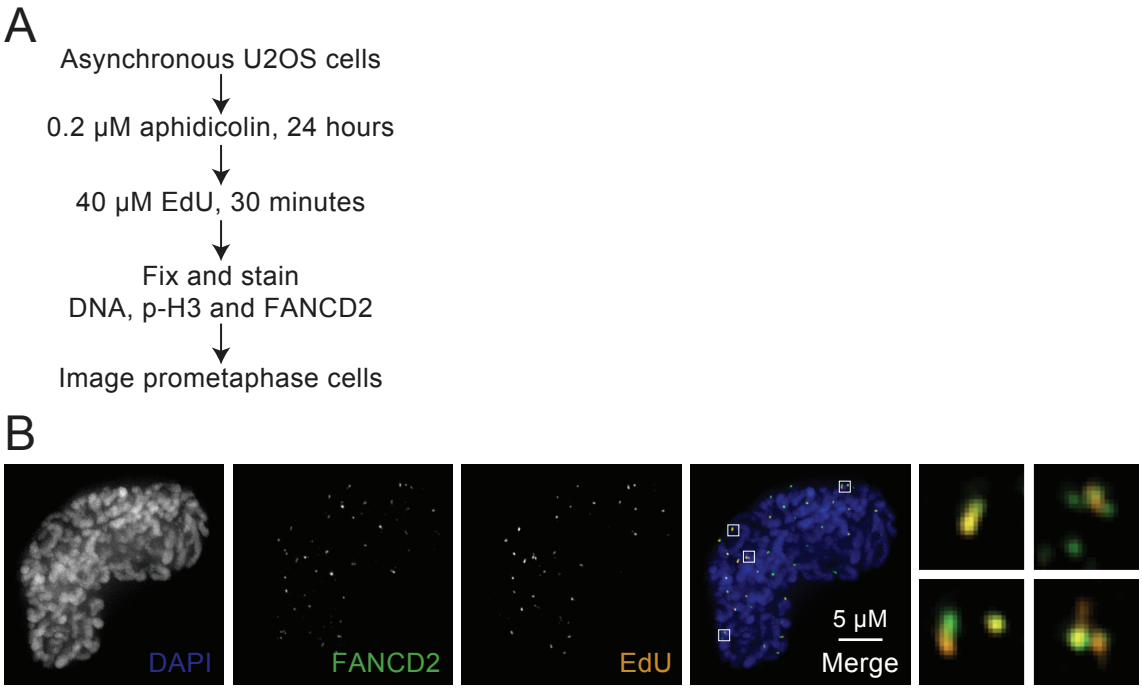


Figure 20: Imaging MiDAS and FANCD2 foci in early-mitotic U2OS cells. A) Schematic of experimental design used in the experiments presented in figures 20, 21 and 22. **B)** Representative image of a pro(meta)phase U2OS cell after a 24 hour treatment with 0.2 μM aphidicolin, stained for FANCD2 (green) and EdU (orange). White squares in 'Merge' pane indicate the context for the inset images.

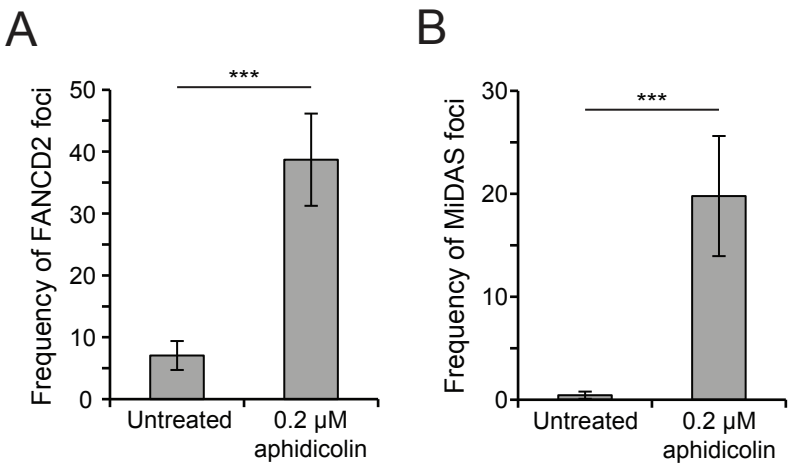


Figure 21: Aphidicolin increases the frequency of FANCD2 and MiDAS foci in early-mitotic cells. A) Quantification of the mean frequency of FANCD2 foci in prometaphase U2OS cells either untreated or incubated with aphidicolin for 24 hours. Error bars represent 95% confidence interval of one biological replicate. t-test, $p = 4.87 \times 10^{-8}$. **B)** Quantification of the mean frequency of MiDAS foci in prometaphase U2OS cells either untreated or incubated with aphidicolin for 24 hours. Error bars represent 95% confidence interval of one biological replicate. t-test, $p = 5.52 \times 10^{-7}$.

prophase and prometaphase U2OS cells there is ~1 MiDAS focus, which increases to ~20 after aphidicolin treatment (figure 21B), which is similar to the quantification in figure 1B of Minocherhomji *et al.* 2015, where RO-3306 was not used either. Encouragingly, a mean frequency of ~1 MiDAS focus is in line with the theoretical prediction for the frequency of DFSs (although the origin number, distribution and genome size in U2OS may differ from HeLa, it is unlikely to differ by an order of magnitude). ~20 MiDAS foci after aphidicolin treatment is similar but slightly higher than the number of G1-specific 53BP1 nuclear bodies quantified in figure 1C of Lukas *et al.* 2011. As inhibiting nucleotide incorporation in mitosis increases the frequency of G1-specific 53BP1 nuclear bodies, and MiDAS foci chased through to G1 infrequently colocalise with 53BP1 nuclear bodies, Minocherhomji *et al.* 2015 propose that when DNA is unsuccessfully repaired by MiDAS that loci is coated in 53BP1 in the following G1 phase. Both frequencies are comparable to the quantification in figure 1B of Minocherhomji *et al.* 2015.

Colocalisation analysis of MiDAS foci and FANCD2 foci is presented in figure 22. In the context of image colocalisation analysis, Pearson's correlation coefficient calculates the regression of a scatter graph of all pixel intensities of one channel against the intensities of the corresponding pixels of the other channel. It can range from -1 (anti-correlative) and 1 (perfect correlation), and works well for foci that are expected to be right on top of each other, and less well for complex structures (such as the inset images in figure 20B) or foci that are merely adjacent or partially overlap. Pearson's correlation coefficient was calculated between MiDAS and FANCD2 foci (figure 22A). Manders' overlap coefficient, designed for the purpose of quantifying colocalisation in biological samples, simply calculates the proportion of background threshold-adjusted voxels of one channel that overlap with the background threshold-adjusted voxels of the other, as well as the reciprocal (for example, there are 4 blue voxels, two of which overlap with red voxels, so Manders' overlap coefficient for the blue channel is 0.5, and for the red channel its 1.0). Manders' overlap coefficient ranges from 0 (not a single voxel overlaps), to 1 (all voxels of one channel overlap with voxels from the other. It is still weak for adjacent foci, but is superior for complex structures and

partially overlapping foci. Manders' overlap coefficients for MiDAS and FANCD2 foci are presented in figure 22B. 'FANCD2<EdU' denotes the proportion of FANCD2 voxels that contain EdU voxels, and 'EdU<FANCD2' is the reciprocal. Finally, MiDAS and FANCD2 foci were defined by an unbiased visual approach and were then judged as either associated or unassociated. The percentage of FANCD2 foci that are associated with MiDAS foci is presented in figure 22C. Only ~5% of FANCD2 foci are associated with MiDAS foci in untreated cells. The average untreated cell has ~5 FANCD2 foci and ~1 MiDAS foci (figure 21), so the maximum expected proportion of overlap is ~20%, and therefore approximately one quarter of MiDAS foci are associated with FANCD2. When treated with aphidicolin, ~50% of FANCD2 foci contain MiDAS foci (figure 22C). The average aphidicolin treated cell contains ~40 FANCD2 foci and ~20 MiDAS foci, so the maximum expected overlap is also ~50%. This is not reflected as a perfect correlation or overlap coefficient in figure 22A and 22B since biological samples are noisy; when two foci colocalise does not necessarily mean every voxel they contain overlap. Overall, treatment with aphidicolin increases the frequencies of MiDAS and FANCD2 foci in early-mitotic U2OS cells (figure 21), but also disproportionately increases the extent of their colocalisation (figure 22). Speculatively, this could be explained by considering colocalised FANCD2 and MiDAS foci as sites of ongoing repair at severe damage. Without perturbation, isolated FANCD2 foci represent its function to protect and restart destabilised replication forks, which occur frequently in unperturbed cells, and MiDAS in isolation are sites of successful gap filling, potentially corresponding to a spontaneous DFS. Aphidicolin increases rates of fork stalling, creating large segments of unreplicated DNA that require BIR-like repair, as proposed by Minocherhomji *et al.* 2015 and Bhowmick *et al.* 2016, with extensive DNA synthesis.

MiDAS foci were further tested for concordance with the DFS model. The experimental design is as before but the 24 hour treatment with aphidicolin is replaced with a 48 hour transfection period with MCM5 siRNA (figure 23A). Figure 23B is a representative image of a prometaphase U2OS cell transfected with control siRNA. MCM5 depletion was verified by immunoblotting, and ~80%

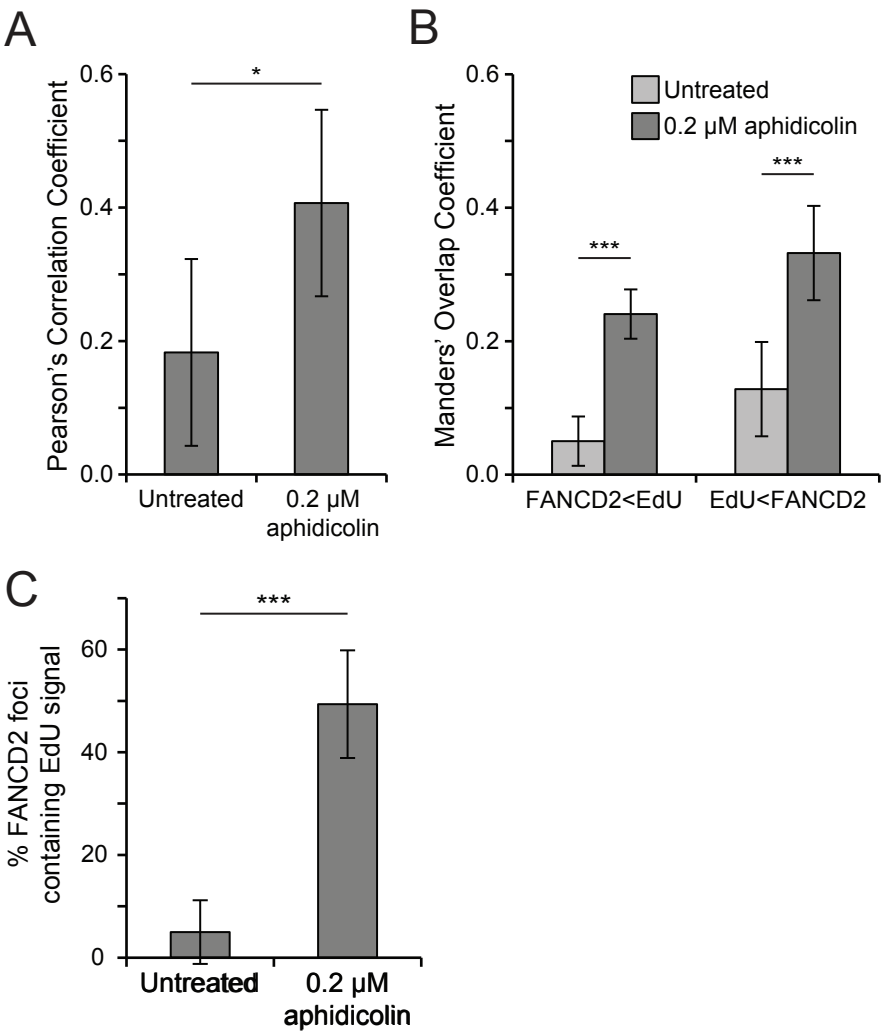


Figure 22: Aphidicolin increases colocalisation of FANCD2 and MiDAS foci. A) Mean Pearson's colocalisation coefficients between FANCD2 foci and MiDAS foci in prometaphase U2OS cells either untreated or incubated with aphidicolin for 24 hours. Error bars represent the 95% confidence interval of one biological replicate. t-test, p = 0.0167. **B)** Mean Manders' overlap coefficient between FANCD2 foci and MiDAS foci in prometaphase U2OS cells either untreated or incubated with aphidicolin for 24 hours. 'FANCD2<EdU' is shorthand for the proportion of FANCD2 signal that overlaps with EdU signal, and the inverse is labelled with 'EdU<FANCD2'. Error bars represent 95% confidence interval of one biological replicate. t-test, $p = 4.06 \times 10^{-6}$, $p = 5.33 \times 10^{-4}$, respectively from left to right. **C)** Quantification of the proportion of FANCD2 foci that contained or overlapped with EdU foci in pro(meta)phase U2OS cells either untreated or incubated with aphidicolin for 24 hours. Error bars represent 95% confidence interval of one biological replicate. t-test, $p = 3.14 \times 10^{-6}$.

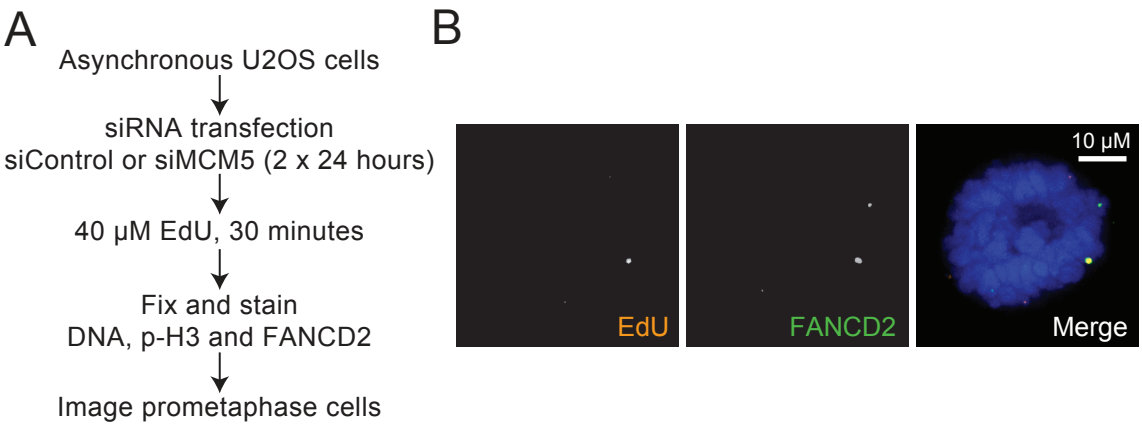


Figure 23: Staining U2OS cells for MiDAS foci and FANCD2 foci after MCM5 RNAi. A) Schematic of experimental design for experiments presented in figures 23, 24 and 25. **B)** Representative image of an unperturbed prometaphase U2OS cell stained with EdU (orange) and FANCD2 (green).

depletion was achieved for each replicate: figure 24A is a representative example. The frequency of FANCD2 foci in early-mitotic cells transfected with either control siRNA or MCM5 siRNA was quantified (figure 24B). The frequency of FANCD2 foci increases from ~3 to ~6.5 after an ~80% depletion of total MCM5. The flow cytometry analysis used previously was not performed on this sample, so the actual reduction in DNA-bound MCM2-7 at S phase entry is unknown. In HeLa cells, an ~80% MCM5 depletion, as measured by immunoblotting whole cell lysate, results in a ~50-60% depletion of DNA-bound MCM2-7 at S phase entry. This reduction of chromatin-bound MCM2-7 is predicted to cause 2-2.5 DFSs per cell, and is associated with a similar number of 53BP1 nuclear bodies in each G1 cell (figure 14). Under these conditions there are twice as many MiDAS foci (~4) and three times as many FANCD2 foci. Most likely, 2-3 FANCD2 foci form for reasons other than a DFS and do not colocalise with MiDAS foci. Alternatively, it could be that FANCD2 foci are the true reflection of DFS frequency, and that G1-specific 53BP1 nuclear bodies are an underestimate. Minocherhomji *et al.* 2015 proposed that unsuccessful repair at MiDAS loci results in the inheritance of the DNA damage which is coated by 53BP1 in the following G1. It is possible that mitotic FANCD2 foci have a similar relationship with 53BP1 nuclear bodies in the following G1. Indeed, depletion of FANCD2 causes an increase in the frequency of cells that contain 53BP1 nuclear bodies (Bourseguin *et al.*, 2016).

Colocalisation analyses were carried out on MiDAS and FANCD2 foci in cells partially depleted of MCM5, as described above. Both Pearson's correlation and Manders' overlap coefficients increase after partial-depletion of MCM5 (figure 25A and 25B). This increase in colocalisation can also be seen in the populations of foci combinations. Foci in each channel were identified with an unbiased approach and were categorised as either 'FANCD2 foci only', 'MiDAS foci only' or 'Colocalised'. Isolated FAND2 foci and MiDAS foci both increase after ~80% MCM5 depletion (figure 25C). However, the greatest change is seen in the colocalised population of foci (figure 25C, yellow columns), and, again, the frequencies of colocalised foci with and without MCM5 partial-depletion are in concordance with the DFS model (figure 2C). As speculated above, specifically colocalised

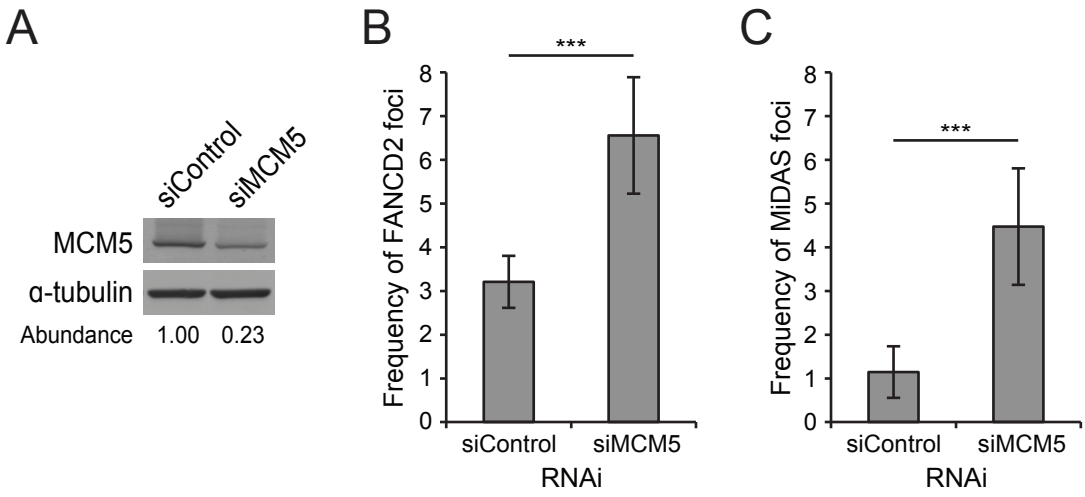


Figure 24: MCM5 RNAi increases the frequency of FANCD2 and MiDAS foci. A) Representative immunoblot of MCM5 depletion in U2OS cells used in the experiments presented in figures 21B and C. Band intensities are normalised against the loading control, α-tubulin. Abundance is band intensity relative to a sample transfected with control siRNA. **B)** Quantification of the mean frequency of FANCD2 foci in prometaphase U2OS cells either transfected with control (scramble) or MCM5 siRNA. Error bars represent 95% confidence interval of three biological experiments. t-test, $p = 2.71 \times 10^{-6}$. **C)** Quantification of the mean frequency of MiDAS foci in prometaphase U2OS cells either transfected with control (scramble) or MCM5 siRNA. Error bars represent 95% confidence interval of three biological experiments. t-test, $p = 3.43 \times 10^{-8}$.

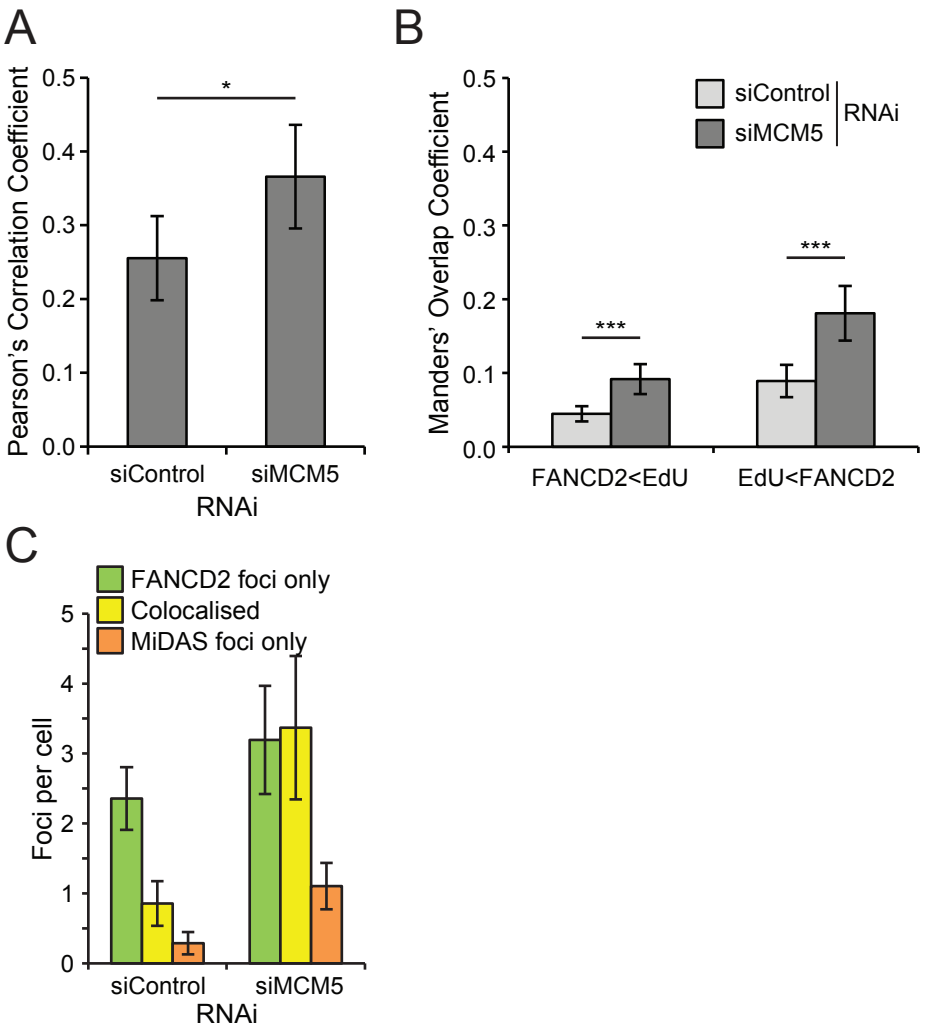


Figure 25: MCM5 RNAi increases colocalisation of FANCD2 and MiDAS foci . A) Mean Pearson's colocalisation coefficients between FANCD2 foci and MiDAS foci in prometaphase U2OS cells either transfected with control (scramble) or MCM5 siRNA. Error bars represent the 95% confidence interval of three biological replicates. t-test, p = 0.0172. **B)** Mean Manders' overlap coefficient between FANCD2 foci and MiDAS foci in prometaphase U2OS cells either transfected with control (scramble) or MCM5 siRNA. 'FANCD2<EdU' is shorthand for the proportion of FANCD2 signal that overlaps with EdU signal, and the inverse is labelled with 'EdU<FANCD2'. Error bars represent 95% confidence interval of three biological experiments. t-test, 2.63×10^{-5}, 2.06×10^{-5}, respectively from left to right. **C)** Quantification of frequency of FANCD2 foci only, MiDAS foci only and where both foci colocalise in prometaphase U2OS cells transfected with either control (scramble) or MCM5 siRNA. Error bars represent 95% confidence interval of three biological experiments. t-test, $p = 1.05 \times 10^{-6}$, for the 'Colocalised' yellow columns.

FANCD2 and MiDAS foci may be sites of a BIR-like DNA synthesis that fills long gaps of unreplicated DNA. Aphidicolin was shown previously to increase MiDAS foci frequency, implicating them as part of CFS repair (Minocherhomji *et al.*, 2015), but here concordance with the DFS model implies that these sites of DNA synthesis are also a response to spontaneous DFSs in unperturbed cells, implicating a specific subset of CFSs that can contain large replicons (e.g. FRA3B and FRA16D).

4.3) UFBs increase in frequency after MCM5 partial-depletion

As mentioned previously, segments of DNA that remain unreplicated up until anaphase will physically link sister-chromatids and cause anaphase bridging, non-disjunction or mis-segregation. To test whether DFSs can cause UFBs, Dr Alberto Moreno imaged UFBs by immunolabelling BLMs helicase, and quantified their frequency in HeLa cells transfected with either control or MCM5 siRNA (figure 26). A representative image is presented in figure 26A, with two UFBs (red) bridging the anaphase gap. The mean frequency of UFBs in control cells is ~1.5, corresponding to the predicted frequency of DFSs and quantified frequency of G1-specific 53BP1 nuclear bodies, as well as previously reported quantifications in unperturbed human cells (Chan *et al.*, 2007, Chan *et al.*, 2009). Here, UFBs were not costained with FANCD2, so Dr Moreno's quantification will include UFBs between centromeres, which are deemed to be of little relevance to CFS repair as they do not increase in frequency after aphidicolin treatment (Chan *et al.*, 2009). It is reported that from ~20% (in unperturbed GM00637 cells) to ~60% (in unperturbed PSNF5 cells, BLM^+ isogen to BLM^- PSNF13) of UFBs are associated with FANCD2 foci (Chan *et al.*, 2009). So, it is likely that Dr Moreno detected just under one UFB in the control transfected HeLa cells. The partial depletion of MCM5 increases the mean frequency of total UFBs to just over 2 (figure 26B). Again, these quantities are in concordance with the predicted frequencies of DFSs with a normal amount of licensing and partially depleted licensing (figure 2C). This claim has been used several times in this thesis so far, so it is worth restating here that it is not just that a cellular event occurs approximately once in an unperturbed cell cycle so must, therefore be a response to DFSs, but that the frequency of

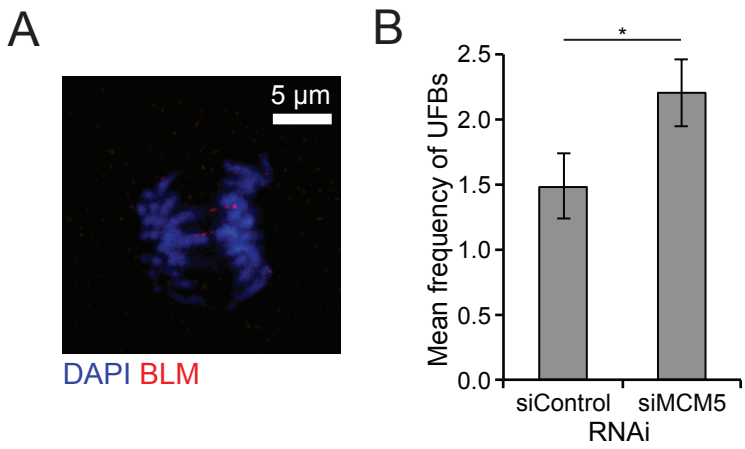


Figure 26: Ultrafine anaphase bridges increase in frequency after MCM5 RNAi. A) Representative image of an anaphase HeLa cell, exhibiting an UFB stained with anti-BLM antibody (red). UFBs imaged by Dr Alberto Moreno. **B**) Quantification of the mean frequency of UFBs in anaphase HeLa cells either transfected with control (scramble) or MCM5 siRNA. Error bars represent S.E.M. of three biological replicates. t-test, p = 0.0473. UFBs quantified by Dr Alberto Moreno.

this event consistently increases to ~2-3 per cell cycle in response to a relatively gentle perturbation that otherwise is undetectable in the absence of induced replicative stress (figure 17 and Ge *et al.* 2007).

4.4) The state of the DNA contained within 53BP1 nuclear bodies inherited by sister-daughter cells

As previously mentioned, there are a few proposed causes of the DNA structures recognised by G1-specific 53BP1 nuclear bodies that are inherited symmetrically by both daughter cells. So far, it has been proposed to be, or a combination of, a) pre-anaphase MUS81 (or other nucleases) nucleolytic cleavage at stalled or collapsed replication forks (Ying *et al.*, 2013), b) DNA breaks caused by tension from chromatin condensation in early-mitosis (Lukas *et al.*, 2011), or c) BLMs helicase unwinding of the unreplicated DNA, producing a segment of ssDNA (Chan and Hickson, 2009, Chan *et al.*, 2007, Harrigan *et al.*, 2011, Lukas *et al.*, 2011).

The DFS model favours the hypothesis of the inheritance of symmetrical ssDNA segments (option c), as it seems likely that this would be the best way to preserve genetic material for both daughter cells. DFSs are predicted to occur approximately once per unperturbed human cell cycle, and unreplicated DNA within a DFS has the potential to be in the order of Mbp in length, so it seems likely that persistent preservation of unreplicated DNA is mandatory for genomic preservation and stability. However, the creation of DSBs by MUS81 or other nucleases (option a) may be a possible consequence of DFSs, particularly for large unreplicated segments, if the unwinding of unreplicated DNA strands is not complete as chromatids are pulled apart in anaphase. We would therefore predict that many 53BP1 nuclear bodies generated by DFS events should contain significant amounts of ssDNA.

To test this hypothesis, HeLa cells were immunofluorescently stained for 53BP1 and the ssDNA-binding protein RPA. As previously reported, ssDNA is not readily detected in G1-specific 53BP1

nuclear bodies (Harrigan *et al.*, 2011, Lukas *et al.*, 2011). Lukas *et al.* 2011, found that <1% of 53BP1 nuclear bodies contained RPA signal, and commented that the little that was found could be explained by the reincorporation of micronuclei, which contain highly damaged 'pulverised' chromatin (Crasta *et al.*, 2012). Although micronucleation can be caused by replication stress or under-replication, reincorporated micronuclei were disregarded from the analysis as they can occur for several reasons, of which a DFS is unlikely. Figure 27A is a representative image of a G1-specific 53BP1 nuclear body that contains what is considered 'legitimate' RPA signal. Examples of micronuclei reincorporation, as stained by 53BP1 and RPA, are in figure 27B, and are distinguished by the distinct lobed or fragmented appearance of the nuclear body. The proportion of G1-specific 53BP1 nuclear bodies that contained 'legitimate' RPA signals was quantified in HeLa cells either transfected with control siRNA or MCM5 siRNA (figure 28). Depletion of MCM5 was confirmed by immunoblot (figure 28A). ~7% of 53BP1 nuclear bodies contained RPA signal (figure 28B). However, partial-depletion of MCM5 increases the proportion of G1-specific nuclear bodies that contain RPA foci to ~30% (figure 28B). The proportion of 53BP1 nuclear bodies colocalised with RPA in control cells is slightly higher here than was reported by Lukas *et. al* 2011, (<1%), and may be explained by sensitivity of the antibodies or microscopes, or the inclusion of false positives, given the subjective distinction between legitimate RPA signal and reincorporated micronuclei. More importantly is that even considering ~7% of 53BP1 nuclear bodies with ssDNA as real is still considerably less than predicted. Harrigan *et al.* 2011 speculate that if 53BP1 do contain significant stretches of ssDNA then it must either be coated in a yet-to-be discovered ssDNA binding protein or has adopted a secondary structure that is not bound by RPA or anti-BrdU antibodies.

A speculative alternative is that MiDAS completes the replication of the vast majority of the unreplicated DNA in mitosis, allowing normal segregation of sister-chromatids, and these sites are then coated in 53BP1 in the following G1, and, like an epigenetic mark, designates the site as previously damaged or in need of further repair. If this is the case then these nuclear bodies would contain little to no ssDNA. If MiDAS partially replicates an unreplicated segment of DNA, for

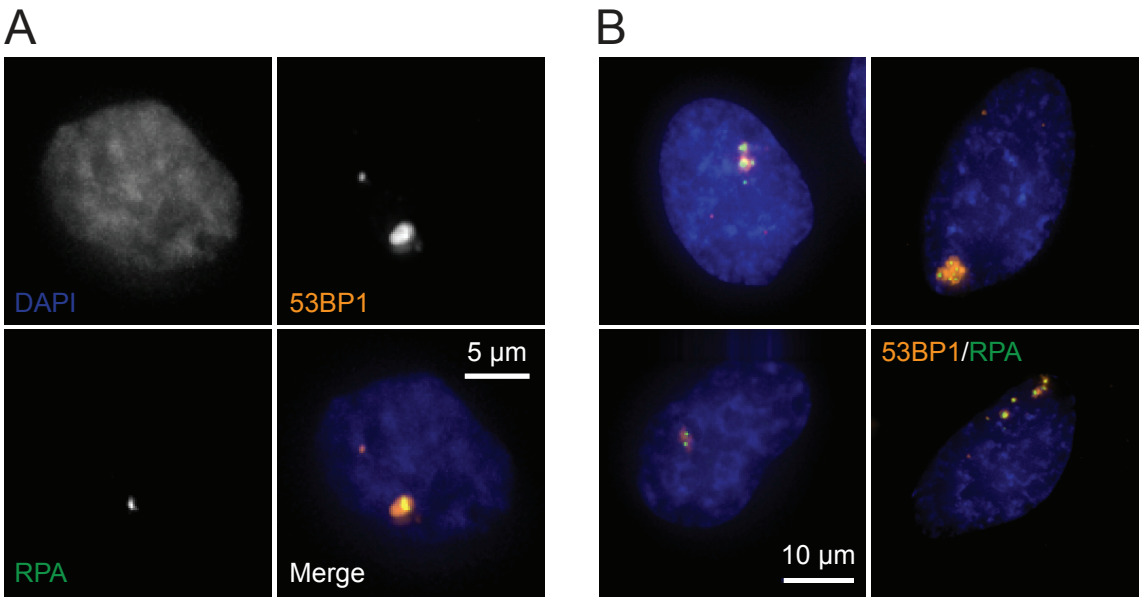


Figure 27: RPA Indicates the Presence of ssDNA in a Proportion of G1-Specific Nuclear Bodies. A) Representative image of an unperturbed HeLa cell stained for 53BP1 (orange) and RPA (green). **B)** Representative images of reincorporated micronuclei, 53BP1 (orange), RPA (green).

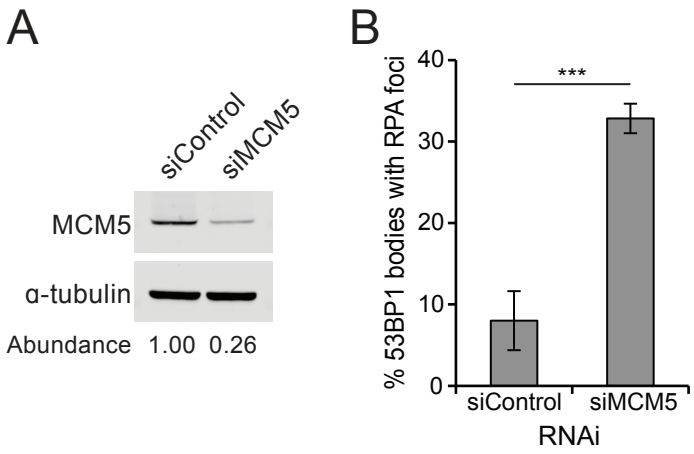


Figure 28: The proportion of G1-specific 53BP1 nuclear bodies that contain RPA signal increases after MCM5 RNAi. A) Representative immunoblot of MCM5 depletion of the HeLa cells used in the experiment presented in figure 28B. Band intensities are normalised against the loading control, α-tubulin. Abundance is band intensity relative to a sample transfected with control siRNA. **B)** Proportion of G1-specific 53BP1 nuclear bodies that contain significant RPA signal in asynchronous HeLa cells either transfected with control (scramble) or MCM5 siRNA. Error bars represent S.E.M. of three biological replicates. t-test, $p = 3.01 \times 10^{-3}$.

example if the unreplicated segment is particularly long, then segregation by BLM helicase or nucleolytic cleavage would still be necessary, and inherited DNA structures would contain partially filled gaps with MiDAS basepairs as well as ssDNA for RPA binding. Partial-depletion of licensing increases the frequency of DFSs, and also their length, which may explain the proportional increase in colocalisation between 53BP1 nuclear bodies and RPA.

To gain a deeper insight into the colocalisation, or lack thereof, between MiDAS foci chased into G1 and G1-specific 53BP1 nuclear bodies, the colocalisation experiments presented in figures 20-25 were used again on the experimental setup described in figure 29A. Asynchronous U2OS cells were incubated with EdU for 30 minutes followed by a mitotic shake-off, washed, re-seeded and allowing to adhere to a cover glass. A representative image of a pair of G1 cells, treated with aphidicolin, is presented in figure 29B. The mean frequency of 53BP1 nuclear bodies in untreated U2OS cells is higher than the control U2OS cells in figure 18D, probably caused by the stress of a mitotic shake-off (figure 30A). Aphidicolin treatment increases the frequency to ~10 nuclear bodies per G1 nucleus (figure 30A). The mean frequency of MiDAS foci in G1 (figure 30B) was ~3, similar to the frequency in early-mitotic cells, as quantified in figure 21. Aphidicolin treatment increases Pearson's correlation and Manders' overlap coefficients for EdU and 53BP1, compared to untreated cells (figure 31A and 31B). ~20% of 53BP1 nuclear bodies contain significant MiDAS foci in untreated cells (see inset images in 29B), which increases to ~50% after aphidicolin treatment (figure 31C). Minocherhomji *et al.* 2015 concluded from a similar experiment that the low level of colocalisation meant that either MiDAS or 53BP1 was sufficient to repair a particular locus of damage, however, they did not compare their colocalisation quantification with untreated cells. The disproportionate increase in colocalisation caused by aphidicolin has similar implications to those from figures 21 and 22. 53BP1 nuclear bodies or MiDAS foci in isolation could represent sites of less severe damage that are occur spontaneously in unperturbed cells, but when cells are treated with aphidicolin both mechanisms are required in the repair of the severe damage caused, and colocalisation in unperturbed cells could be the response to a DFS.

As mentioned above, 53BP1 nuclear bodies could reflect dsDNA breaks, either generated as a consequence of DFSs or by other means. The frequency of G1-specific ɣ-H2AX foci was quantified to test if under-replication would increase DSBs as detected by ɣ-H2AX in the following G1. As with figure 3A, G1 cells were identified in an asynchronous population by the absence of EdU and CyclinA signal (figure 32A). ɣ-H2AX foci are frequent in S phase (figure 32A, top left cell), and localise to sites of replication stress, and are disregarded. HeLa cells were partially depleted of MCM5 (figure 32B). There is no significant difference in frequency between G1 HeLa cells transfected with control siRNA or MCM5 siRNA (figure 32C). A more comprehensive colocalisation analysis with 53BP1 would be required to fully characterise the DNA content of 53BP1 nuclear bodies in response to partial depletion of licensing, but this initial result indicates that there might be distinct pathways used by most CFS resolution and most DFS resolution. There are multiple ways in which CFSs are expressed by aphidicolin treatment that could develop into DSBs, perhaps a partial-depletion of licensing specifically exacerbates those CFSs that are contained within large replicons, and are less likely to produce DSBs come G1 phase.

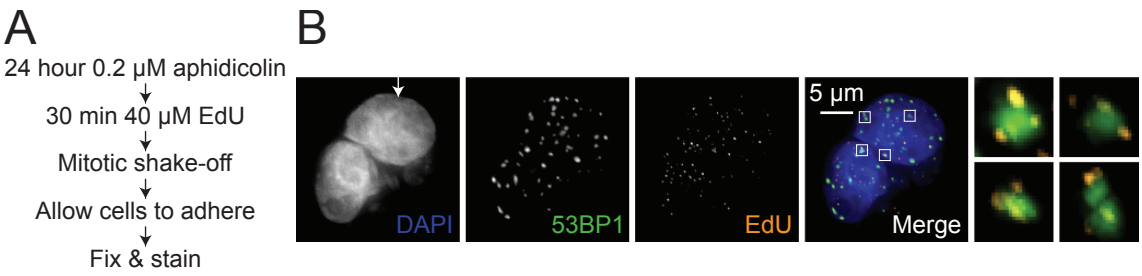


Figure 29: MiDAS foci from Early-mitosis are localised within G1-specific 53BP1 nuclear bodies. A) Schematic of experimental design to observe the fate of pro(meta)phase MiDAS foci in G1 cells immunostained for 53BP1. **B)** Representative image of G1 U2OS cells treated with aphidicolin for 24 hours immunostained for 53BP1 (green) and EdU incorporated in early mitosis (orange).

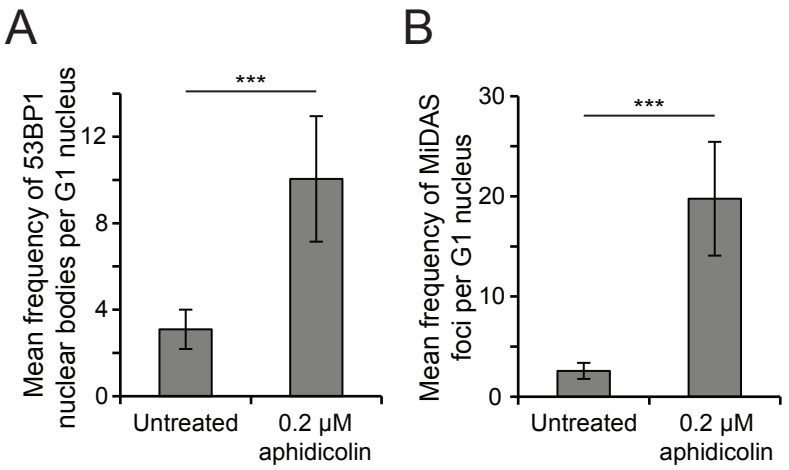


Figure 30: MiDAS foci and 53BP1 nuclear bodies both increase in frequency in G1 cells after treatment with aphidicolin. A) Quantification of the mean frequency of G1-specific 53BP1 nuclear in U2OS cells either untreated or incubated with aphidicolin for 24 hours. Error bars represent 95% confidence interval of one biological replicate. t-test, $p = 6.07 \times 10^{-5}$. **B)** Quantification of the mean frequency of MiDAS foci from pro(meta)phase in G1 U2OS cells either untreated or incubated with aphidicolin for 24 hours. Error bars represent 95% confidence interval of one biological replicate. t-test, $p = 7.07 \times 10^{-7}$.

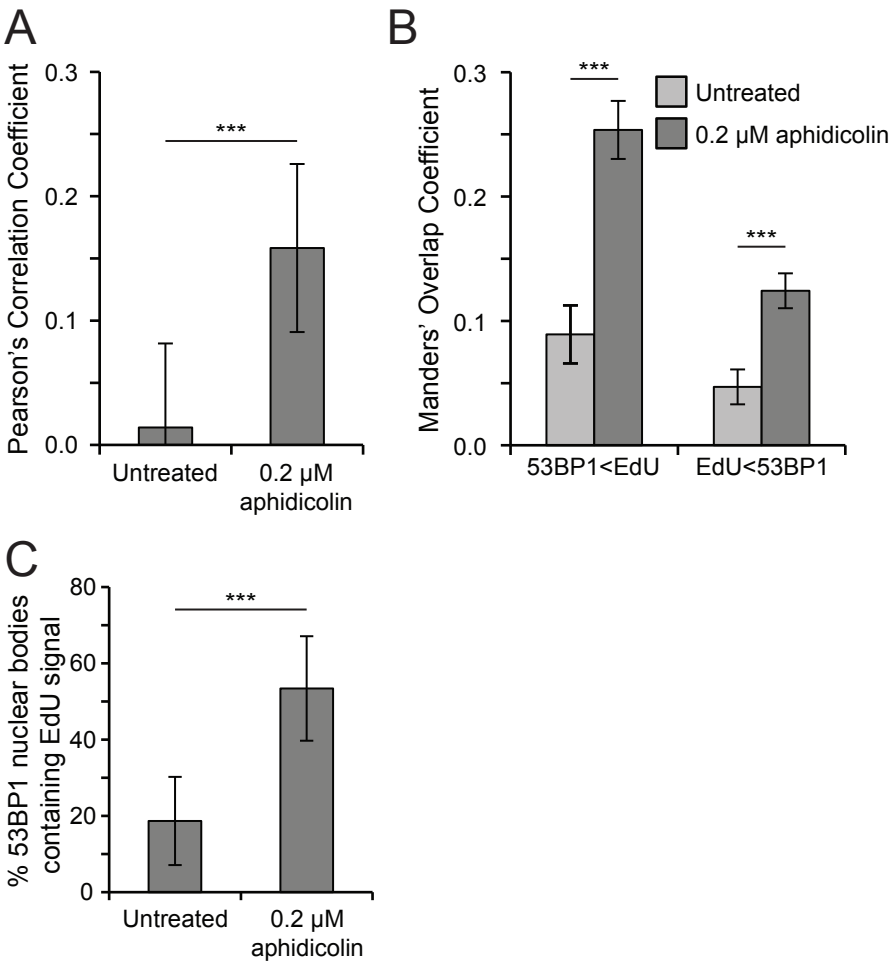


Figure 31: MiDAS foci increase in colocalisation with G1-specific 53BP1 nuclear bodies after treatment with aphidicolin. A) Quantification of the proportion of G1-specific 53BP1 nuclear bodies that contained or overlapped with MiDAS foci incorporated in pro(meta)phase U2OS cells either untreated or incubated with aphidicolin for 24 hours. Error bars represent 95% confidence interval of one biological replicate. t-test, $p = 2.80 \times 10^{-3}$. **B)** Mean Pearson's colocalisation coefficient between G1-specific 53BP1 nuclear bodies and MiDAS foci from pro(meta)phase in G1 U2OS cells either untreated or incubated with aphidicolin for 24 hours. Error bars represent he95% confidence interval of one biological replicate. t-test, $p = 2.70 \times 10^{-3}$. **C)** Mean Manders' overlap coefficient between G1-specific 53BP1 nuclear bodies and MiDAS foci from pro(meta)phase in G1 U2OS cells either untreated or incubated with aphidicolin for 24 hours. '53BP1<EdU' is shorthand for the proportion of 53BP1 signal that overlaps with EdU signal, and the inverse is labelled with 'EdU<53BP1'. Error bars represent 95% confidence interval of one biological replicate. t-test, $p = 9.51 \times 10^{-8}$, $p = 5.55 \times 10^{-4}$, respectively from left to right.

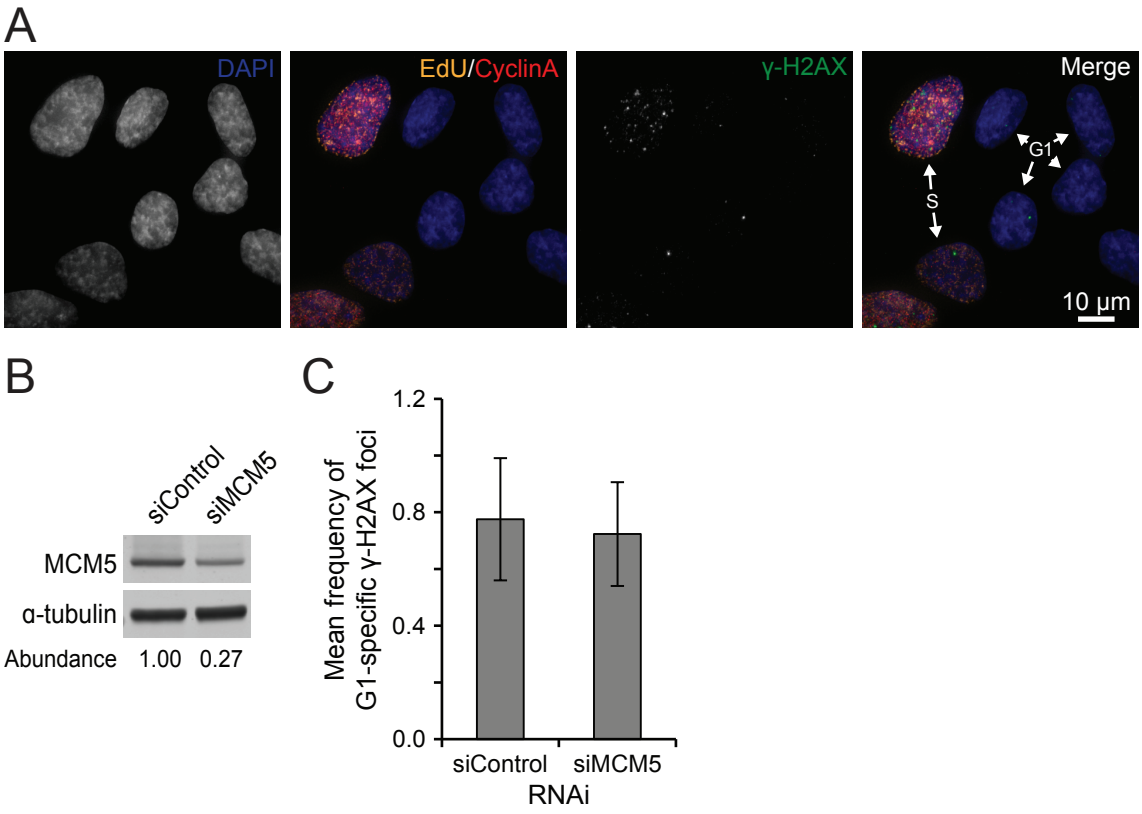


Figure 32: MCM5 RNAi does not significantly increase G1-specific double strand breaks as marked by γ-H2AX foci. A) Representative of asynchronous HeLa cells stained with γ-H2AX (green), EdU (orange) and CyclinA (red). EdU and CyclinA staining are used to stage cells (white arrows) and G1 cells are used for the quantification in figure 32C. **B)** Representative immunoblot of the HeLa cells used in the experiment presented in figure 32C. Band intensities are normalised against the loading control, α-tubulin. Abundance is band intensity relative to a sample transfected with control siRNA. **C)** Quantification of the mean frequency of G1-specific γ-H2AX in asynchronous HeLa cells transfected with either control (scramble) or MCM5 siRNA. Error bars represent 95% confidence interval of three biological replicates. t-test, p = 0.716.

Chapter 5

Discussion

5.1) Summary of results

Presented in this thesis is evidence that DFSs occur at a significantly higher rate in unperturbed human cells compared to yeasts, with which the DFS model was originally validated (Newman *et al.*, 2013), and that the cellular responses to DFSs overlap with the cellular responses to CFSs. Firstly, given previous investigations and speculation as to the function of G1-sepcific 53BP1 nuclear bodies (Harrigan *et al.*, 2011, Lukas *et al.*, 2011), they were tested as a candidate for responding to DFSs. The frequency of 53BP1 nuclear bodies in unperturbed cells, and in cells with partially depleted licensing, is in concordance with the predicted frequency of DFSs, and is likely applicable to primary cell lines. 53BP1 preferentially binds to chromatin associated with large replicons in unperturbed cells, and functions with dormant origins to protect genome stability against replication stress. Secondly, research into cellular responses to CFS expression has identified a series of factors that facilitates recovery from replication stress after the completion of S phase, and of those examined, were also found to largely concord with the predicted frequency of DFSs in unperturbed cells and cells partially depleted of licensing. These factors include FANCD2 and MiDAS, as well as BLM-coated UFBs. Thirdly, after partial licensing depletion, the low frequency of G1-specific 53BP1 nuclear bodies that contain RPA (ssDNA) and MiDAS foci increases whilst the higher frequency that contain ɣ-H2AX (dsDNA) do not.

This work builds on the original DFS model (Newman *et al.*, 2013), and indicates that it is also applicable to higher eukaryotes, and therefore a conserved side-effect of the eukaryotic licensing system. Additionally, it is likely that cellular responses to DFSs are the same as those for some CFSs, which is consistent with the subset of CFSs that are associated with very large replicons, such as FRA3B and FRA16D (Letessier *et al.*, 2011). It is also likely a significant proportion of DFSs occur at CFSs in unperturbed cells, but could also potentially occur anywhere in the genome.

5.2) Model

Collating the results presented here, as well as extensive research into CFSs, a model for the cellular responses to DFSs in unperturbed cells is presented in figure 33. From top to bottom is progression through the cell cycle from an arbitrary G1 to the S phase of the following cell cycle (stages are indicated in the left-hand margin). Starting with a duplex of unreplicated G1 DNA (blue), one or more DFSs occur in ~70% of all S phases. Theoretical modelling of the frequency distribution of DFSs in human cell S phases predicts that ~30% of S phases exhibit zero DFSs (see figure 6 of Al Mamun *et al.* 2016), which is reflected in the frequency distribution of G1-specific 53BP1 nuclear bodies in HeLa cells (figure 4B), where ~40% of G1 cells contain no 53BP1 nuclear bodies (figure 33, ⓐ).

A mean of ~1-2 DFSs are predicted per HeLa S phase (figure 2C and 4A). Mechanisms that can facilitate fork restart are acknowledged here, and can facilitate the complete duplication of the genome in response to under-replication (figure 33, ⓐ). For example, recent work published by Natsume *et al.* 2017, revealed the MCM8/9 can support progress through S phase via HR after induction of acute MCM2 destruction during S phase (Natsume *et al.*, 2017). Similarly, the activity of MUS81 complexed with EME2 is restricted to S phase and promotes fork restart (Pepe and West, 2014). However, one of the primary assumptions of the DFS model is that there is a per nucleotide probably of irreversible fork stalling, which is built into the estimation for N_s (Newman *et al.*, 2013), so fork restart or HR-mediated DNA synthesis that compensate for fork stalling in general are disregarded.

FANCD2 is recruited to stalled forks, to stabilise and potentially facilitate restart (figure 33, ⓐ). However, where fork restart fails, MUS81-EME1 is recruited to stalled forks in mitosis and cleaves them to repress formation of UFBs. There is a mean frequency of ~3-5 FANCD2 foci per early-mitotic U2OS cell (figure 21A and 24B), which is approximately double the predicted frequency of

DFSs (and UFBs, and 53BP1 nuclear bodies). This might indicate a function of these foci independent of DFSs, or might reflect the structure of a DFS, which is composed of two stalled forks but only causes one UFB and one 53BP1 nuclear body per sister-G1 cell (so two nuclear bodies total). Successful cleavage of stalled forks by MUS81-EME1 may result in the loss of the unreplicated segment of DNA, which is perhaps preferable to risking mis-segregation and UFBs. Each daughter cell inherits the DSB ends, which are maintained in local association by MRN scaffolding function, and localised activation of ɣ-H2AX promotes the recruitment of 53BP1 to chromatin ~1 Mbp around the DSB, and carries out its well described role in promoting NHEJ (figure 33, ⓑ). Chromosome breaks are locally enriched with MDC1 (a direct sensor of ɣ-H2AX), a well characterised DSB marker, which persist from at least metaphase through to G1 and form symmetrical nuclear bodies in the daughter cells that colocalise with 53BP1. Additionally, ɣ-H2AX foci are symmetrical in number between segregated anaphase chromosomes and both of these phenomena occur at a low frequency in unperturbed cells, which increases after aphidicolin treatment (Lukas *et al.*, 2011). This indicates that MUS81-type nuclease cleavage, or tension from chromosome condensation, creates DSBs in early-mitosis and account for a proportion of G1-specific 53BP1 nuclear bodies. The function of these G1-specific 53BP1 nuclear bodies would then be in concordance with the genetic studies that identified 53BP1 as an inhibitor of HR and DNA end resection and a facilitator of NHEJ (Bothmer *et al.*, 2010, Bouwman *et al.*, 2010, Bunting *et al.*, 2010).

In an alternative pathway, but not necessarily mutually exclusive, MUS81 cleavage could facilitate MiDAS upon chromosome-condensation in early-mitosis. The DSB ends created by MUS81-mediated cleavage are resected and coated in RAD52 to facilitate microhomology-mediated BIR (Bhowmick *et al.*, 2016), which results in the completion of replication at the previously unreplicated segment of DNA (figure 33, ⓒ). The mean frequency of MiDAS foci in unperturbed early-mitotic U2OS cells is ~1 (figure 21B and 24C). BIR synthesis is highly inaccurate (Deem *et al.*, 2011) which potentially explains why some segments of MiDAS are coated in 53BP1,

preventing end resection and HR, transcription (Harrigan *et al.*, 2011), and protecting the mutagenised DNA until high-fidelity HR-dependant DNA repair can take place in S phase. However, in contradiction to this, BIR-type DNA synthesis is conservative (Donnianni and Symington, 2013, Saini *et al.*, 2013), so to preserve genetic information one daughter cell would inherit a MiDAS duplex and the other the original unreplicated segment of DNA (figure 33, ©, top). Bhowmick *et al.* 2016, in the follow up to the model they presented in Minocherhomji *et al.* 2015, prepared metaphase spreads and visualised MiDAS foci. Quantification revealed that ~45% of MiDAS foci were situated on one chromatid arm only, and ~35% were situated on both chromatid arms (~20% were uncategorisable), implying that there is more than one MiDAS mechanism or indeed that RO-3306 synchronisation and release is imperfect, as mentioned in the introduction, and semiconservative S/G2 synthesis has been visualised. It is not clear how an asymmetric process such as this would cause symmetrical FANCD2 foci in telophase (Chan *et al.*, 2009) and symmetrical 53BP1 nuclear bodies in the following G1 (Lukas *et al.*, 2011). Perhaps both the 'broken' and 'donor' chromosomes require ongoing repair following BIR, and would explain why not all 53BP1 nuclear bodies contain MiDAS foci, as one daughter cell would inherit a MiDAS duplex and coat it in 53BP1, and the other daughter cell would inherit the original unreplicated segment of DNA and coat it in 53BP1 as well. If MiDAS is BIR-like (conservative) and does not complete gap filling then one daughter would inherit the original unreplicated segment of DNA and the other a deletion that can potentially span the length of the unreplicated segment. In the context of the DFS model, the potential for deletions such as this occur a mean of 1-2 times per cell cycle, and would resemble genetic loss reminiscent of telomere shortening, but from the body of chromosomes. Alternatively, MiDAS could be semi-conservative (figure 33, ©, bottom), consistent with the symmetric telomeric FANCD2 foci (Chan *et al.*, 2009) and symmetrical 53BP1 nuclear bodies in the following G1 (Lukas *et al.*, 2011). Semi-conservative MiDAS would likely result in less genetic loss when compared to the conservative version, and so is a satisfying solution to the predicted high frequency of DFSs.

If the unreplicated segment is not cleaved by MUS81-EME1, or replicated by MiDAS, then sister-chromatids will remain physically linked during chromosome segregation in anaphase, forming a FANCD2-flanked and BLM-coated UFB. Dr Alberto Moreno quantified the frequency of UFBs in unperturbed HeLa cells to be ~1-2 per anaphase cell, in agreement with the predicted frequency of DFSs. If BLM helicase and topoisomerase IIIα activities successfully unwind the unreplicated segment of DNA then each daughter cell inherits a symmetrical segment of ssDNA, which could potentially be coated by 53BP1 and protected from HR in G1 and repaired by high-fidelity HR during S phase (figure 33, ⓓ). Like semi-conservative MiDAS, this mechanism would likely stably maintain genetic information, a requirement that the high frequency of DFSs would necessitate. It would also explain why 53BP1 nuclear bodies are at a constant frequency through G1 and decline during S phase (figure 6 and 7), which ⓑ does not. If semi-conservative MiDAS only partially replicates previously unreplicated DNA then symmetrical 53BP1 nuclear bodies in daughter cells would be expected to contain a low frequency of MiDAS foci and RPA. ~20% of G1-specific 53BP1 nuclear bodies contain MiDAS foci (figure 31C) and ~7% contain RPA (figure 28B). Both frequencies increase after aphidicolin treatment and partial licensing depletion respectively, possibly because the average length of unreplicated segments has increased, making merely partial gap-filling by MiDAS more likely.

Overall, these results are consistent with the idea that there are multiple mechanisms that function to resolve under-replication at specific cell cycle stages, starting with fork stabilisation and restart in S phase, through to the accumulation and disappearance of 53BP1 nuclear bodies in the following S phase. Each mechanism can feasibly resolve un-replicated segments of DNA, and given the order in which they function, are reminiscent of layered safety nets under a trapeze gymnast: if fork restart fails, then MUS81 is recruited to repress UFB formation and/or facilitate MiDAS; if MiDAS fails, then UFBs form and BLM helicase is recruited to unwind the DNA. Whether daughter cells inherit DSBs from MUS81-mediated cleavage, ssDNA segments from BLM helicase activity, and/or nucleotides synthesised by MiDAS, then 53BP1 is recruited to those lesions to protect against

inappropriate HR until high-fidelity repair in S phase. Given the predicted frequency of ~1-2 DFSs per S phase, it would make sense that a requirement of their resolution would be the preservation of the genetic information contained within the unreplicated DNA. If so, it would seem semi-conservative MiDAS and/or the pathway that leads to ⓓ should be favoured over the potentially deleterious conservative version of MiDAS or the pathway that leads to ⓑ.

5.3) G1-sepcific 53BP1 nuclear bodies in primary and cancer human cell lines

Despite the similar genome size, origin frequency and origin distribution between HeLa and IMR-90 cells, HeLa cells exhibit approximately double the number of 53BP1 nuclear bodies as IMR-90 cells (figure 4A and figure 19D) (Al Mamun *et al.*, 2016, Picard *et al.*, 2014). It is possible that the difference between the untransformed cell line, IMR-90, and the transformed cancer cell line, HeLa, is ploidy. The telomere-to-genome size ratio is greater in HeLa cells, which will have ~150 telomeres in G1 and ~300 by G2 and M, compared to IMR-90 which will have ~100 in G1 and ~200 in G2 and M. As discussed and modelled in Newman *et al.* 2013, telomeres are unusual in that if the last replication fork that comes from the body of the chromosome stalls then it is equivalent to a DFS event, in that a segment of DNA remains unreplicated, but are more likely to occur as only a single fork must stall. Additionally, telomeres are G-rich, contain repetitive sequences, are coated in proteins and form secondary structures, all of which can contribute to fork stalling (Webb *et al.*, 2013). This hypothesis was validated in yeasts by how non-randomly close the last origins are to the telomere, an average of ~0.4 kb, compared to the average replicon length of ~26 kbp (Newman *et al.*, 2013). Additionally, Theis *et al.* 2010 compared the loss rate of a 160 kbp origin-less segment of DNA in the body of a chromosome to the same segment at the end of a chromosome, and found that there was a ~200-fold increase. The DFS model predicted a ~1000-fold increase and Newman *et al.* 2013 speculated whether there is a mechanism to resolve under-replication at telomeres. Indeed, telomerase-null *S. cerevisiae* cells mostly die but the few that survive use an ALT pathway that involves recombination between uneven telomeres to maintain their length (Teng and Zakian, 1999). BLM helicase is required for ALT in human cells (Stavropoulos *et al.*, 2002), localises to

telomeres, and depletion by RNAi causes an increase in telomere free ends (Barefield and Karlseder, 2012). Inducing ALT by inhibition of TRF2 (telomere repeat-binding factor, a member of the shelterin complex that protects telomeres) causes an increase in 53BP1 nuclear bodies in the following G1, and repair factors including 53BP1 colocalise with unprotected telomeres (Takai *et al.*, 2003). It is possible that the ~1.5 fold increase in telomeres in HeLa cells compared to IMR-90 are creating more difficult to replicate regions of DNA that infrequently require 53BP1 localisation.

Alternatively, the reduced frequency of 53BP1 nuclear bodies could be explained by the presence of an intact licensing checkpoint in IMR-90 cells, that ensures a sufficient complement of licensed replication origins before cells enter S phase. However, this is not reflected in the origin-mapping data of IMR-90, where HeLa and IMR-90 are detected as having a similar genome size and number of origins (Al Mamun *et al.*, 2016, Picard *et al.*, 2014), which is reflected in the predicted frequencies of DFSs (figure 2C, the difference that is there is largely in the coefficient of variation, *R*, which is slightly smaller in IMR-90 cells).

Finally, the reduced frequency of 53BP1 nuclear bodies in IMR-90 cells could be explained by the presence of a well-regulated DNA damage response. In normal cells, DNA damage and replication stress are detected by the DNA damage response network by the central regulator kinases ATM and ATR respectively. ATM and ATR transduce their signals via p53 which can then coordinate apoptosis, senescence or cell cycle arrest (among other activities) (Marechal and Zou, 2013). In the absence of p53 activity, a hallmark of many cancers, cells proliferate uncontrollably. A sufficient number of DFSs in an IMR-90 S phase could trigger ATR- and p53-mediated cell cycle arrest (and restarted again after resolution) or senescence, causing an underestimate of DFSs as detected by 53BP1 nuclear bodies.

Either way, the increased reliance on resolution of under-replication in cancer cells could be exploited as an anti-cancer drug target. Indeed, recently published work from the Blow lab indicates

that cancer cells are hypersensitive to the partial inhibition of licensing (Gardner *et al.*, 2017). A library of small molecules was screened for activity that reduces the amount of chromatin-bound MCMs in U2OS cells. A class of 2-arylquinolin-4-amines that disrupted ORC binding to DNA in both *Xenopus* cell-free extract and human cells was discovered. U2OS proliferation is hypersensitive to treatment with the most potent licensing inhibitor discovered by the screen, when compared to the same treatment of the primary cell line IMR-90 (Gardner *et al.*, 2017). U2OS cells lack an intact licensing checkpoint and likely entered S phase sub-minimally licensed and greatly destabilised their genomes, whereas the licensing checkpoint in IMR-90 cells would merely have stalled them in G1. Cycling drug treatments could potentially stop host cells from senescing whilst remaining lethal to the cancer cells. Combining the treatment of the licensing inhibitor with something that inhibits resolution of under-replication could be significantly more lethal to cancer cells (for example, figure 17B), whilst little more lethal to host cells as they are blocked from entering S phase.

5.4) Implications of the DFS model on genome size and origin distribution

Increasing genome size from megabases in yeasts to gigabases in metazoans causes the probability of one or more DFSs to increase from ~0.1% to ~70% of S phases (Al Mamun *et al.*, 2016). In yeasts, optimising the distribution of origins still has a significant effect on proportionately reducing the predicted frequency of DFSs (Newman *et al.*, 2013), but for human cells, evenly spacing origins would not compensate for the huge increase in genome size (Al Mamun *et al.*, 2016). Al Mamun *et al.* 2016 hypothesise that mechanisms that resolve DFSs lifts the burden of optimising origin distribution in higher-eukaryotes.

Speculatively, this could explain why metazoans have lost any recognisable origin consensus sequence, and why the *S. cerevisiae* genome has lost most of its introns. Regulating the positions of ARSs could be how yeasts ensure the relatively even spacing between their origins, so the absence of metazoan ARSs could be explained by the loss of the benefit of even spacing in gigabase

genomes with under-replication resolution mechanisms, and why human ORC is primary DNA sequence-agnostic (Schaarschmidt *et al.*, 2004, Vashee *et al.*, 2003). The *S. cerevisiae* genome is known to have few introns, having lost them compared to an intron-rich fungal ancestor (Hooks *et al.*, 2014). Budding yeasts that spontaneously lost introns by transposon or recombination events may have had an advantage over other yeasts by further optimising their genome to reduce DFSs.

From these considerations, it seems clear that one of the major requirements for expanding the size of a genome from megabases to gigabases are mechanisms that can resolve spontaneously occurring under-replication.

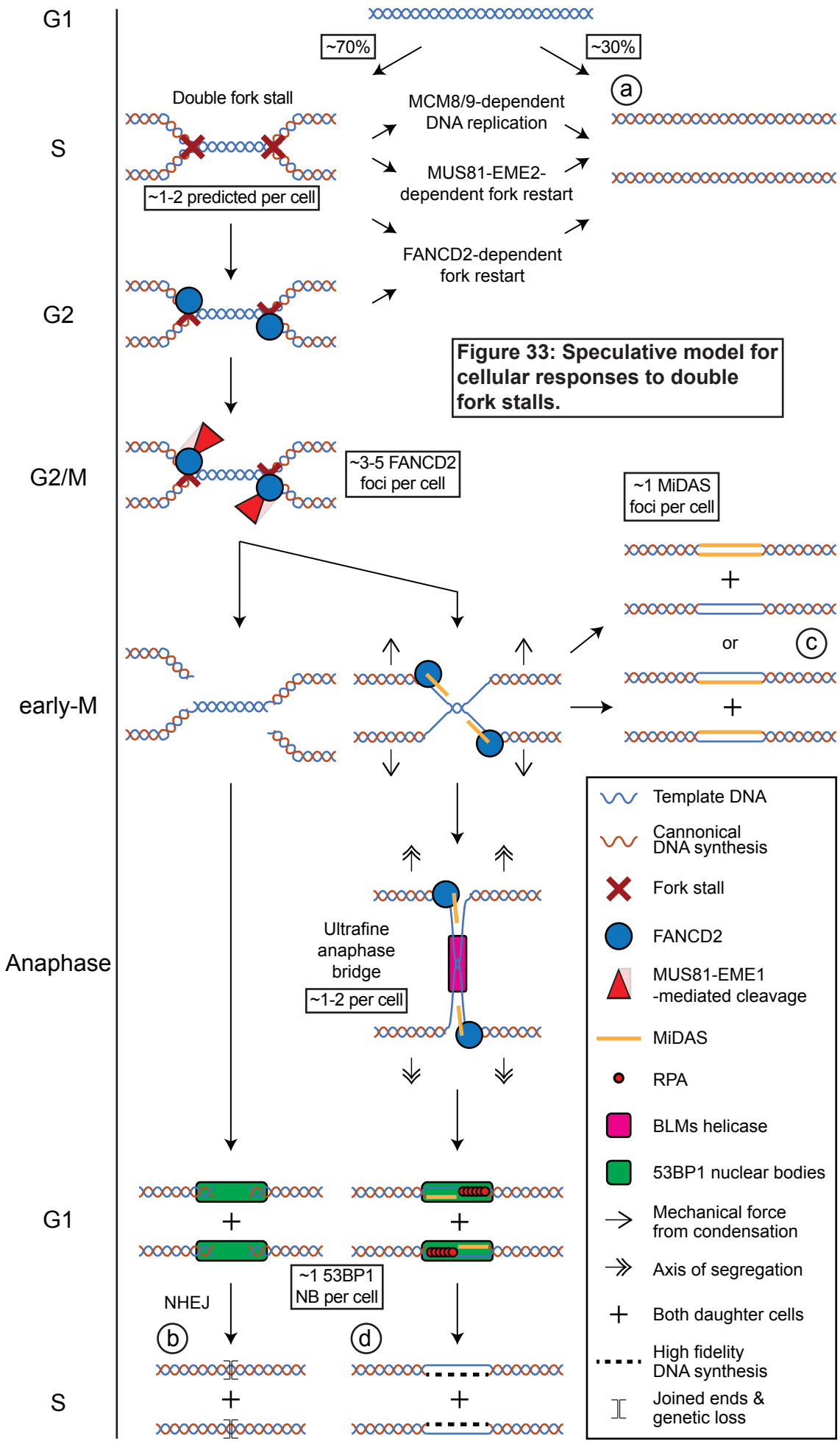


Figure 33: Speculative model for cellular responses to double fork stalls.

Bibliography

Al Mamun, M., Albergante, L., Moreno, A., Carrington, J. T., Blow, J. J. & Newman, T. J. 2016. Inevitability And Containment Of Replication Errors For Eukaryotic Genome Lengths Spanning Megabase To Gigabase. *Proceedings Of The National Academy Of Sciences Of The United States Of America,* 113**,** E5765-74.

Alabert, C. & Groth, A. 2012. Chromatin Replication And Epigenome Maintenance. *Nat Rev Mol Cell Biol,* 13**,** 153-67.

Aladjem, M. I. 2004. The Mammalian Beta Globin Origin Of DNA Replication. *Front Biosci,* 9**,** 2540-7.

Alver, R. C., Chadha, G. S. & Blow, J. J. 2014. The Contribution Of Dormant Origins To Genome Stability: From Cell Biology To Human Genetics. *DNA Repair (Amst),* 19**,** 182-9.

Anderson, L., Henderson, C. & Adachi, Y. 2001. Phosphorylation And Rapid Relocalization Of 53bp1 To Nuclear Foci Upon DNA Damage. *Mol Cell Biol,* 21**,** 1719-29.

Arias, E. E. & Walter, J. C. 2007. Strength In Numbers: Preventing Rereplication Via Multiple Mechanisms In Eukaryotic Cells. *Genes Dev,* 21**,** 497-518.

Aula, P. & Von Koskull, H. 1976. Distribution Of Spontaneous Chromosome Breaks In Human Chromosomes. *Hum Genet,* 32**,** 143-8.

Bachrati, C. Z. & Hickson, I. D. 2009. Dissolution Of Double Holliday Junctions By The Concerted Action Of Blm And Topoisomerase Iiialpha. *Methods Mol Biol,* 582**,** 91-102.

Barefield, C. & Karlseder, J. 2012. The Blm Helicase Contributes To Telomere Maintenance Through Processing Of Late-Replicating Intermediate Structures. *Nucleic Acids Res,* 40**,** 7358-67.

Bartkova, J., Rezaei, N., Liontos, M., Karakaidos, P., Kletsas, D., Issaeva, N., Vassiliou, L. V., Kolettas, E., Niforou, K., Zoumpourlis, V. C., Takaoka, M., Nakagawa, H., Tort, F., Fugger, K., Johansson, F., Sehested, M., Andersen, C. L., Dyrskjot, L., Orntoft, T., Lukas,

J., Kittas, C., Helleday, T., Halazonetis, T. D., Bartek, J. & Gorgoulis, V. G. 2006. Oncogene-Induced Senescence Is Part Of The Tumorigenesis Barrier Imposed By DNA Damage Checkpoints. *Nature,* 444**,** 633-7.

Bassing, C. H. & Alt, F. W. 2004. H2ax May Function As An Anchor To Hold Broken Chromosomal DNA Ends In Close Proximity. *Cell Cycle,* 3**,** 149-53.

Bastin-Shanower, S. A., Fricke, W. M., Mullen, J. R. & Brill, S. J. 2003. The Mechanism Of Mus81-Mms4 Cleavage Site Selection Distinguishes It From The Homologous Endonuclease Rad1-Rad10. *Mol Cell Biol,* 23**,** 3487-96.

Baumann, C., Korner, R., Hofmann, K. & Nigg, E. A. 2007. Pich, A Centromere-Associated Snf2 Family Atpase, Is Regulated By Plk1 And Required For The Spindle Checkpoint. *Cell,* 128**,** 101-14.

Bekker-Jensen, S., Lukas, C., Melander, F., Bartek, J. & Lukas, J. 2005. Dynamic Assembly And Sustained Retention Of 53bp1 At The Sites Of DNA Damage Are Controlled By Mdc1/Nfbd1. *J Cell Biol,* 170**,** 201-11.

Bell, S. P. 1995. Eukaryotic Replicators And Associated Protein Complexes. *Curr Opin Genet Dev,* 5**,** 162-7.

Bell, S. P. & Dutta, A. 2002. DNA Replication In Eukaryotic Cells. *Annu Rev Biochem,* 71**,** 333-74.

Berezney, R., Dubey, D. D. & Huberman, J. A. 2000. Heterogeneity Of Eukaryotic Replicons, Replicon Clusters, And Replication Foci. *Chromosoma,* 108**,** 471-84.

Berti, M., Ray Chaudhuri, A., Thangavel, S., Gomathinayagam, S., Kenig, S., Vujanovic, M., Odreman, F., Glatter, T., Graziano, S., Mendoza-Maldonado, R., Marino, F., Lucic, B., Biasin, V., Gstaiger, M., Aebersold, R., Sidorova, J. M., Monnat, R. J., Jr., Lopes, M. & Vindigni, A. 2013. Human Recq1 Promotes Restart Of Replication Forks Reversed By DNA Topoisomerase I Inhibition. *Nat Struct Mol Biol,* 20**,** 347-54.

Berti, M. & Vindigni, A. 2016. Replication Stress: Getting Back On Track. *Nat Struct Mol Biol,* 23**,** 103-9.

Besnard, E., Babled, A., Lapasset, L., Milhavet, O., Parrinello, H., Dantec, C., Marin, J. M. & Lemaitre, J. M. 2012. Unraveling Cell Type-Specific And Reprogrammable Human Replication Origin Signatures Associated With G-Quadruplex Consensus Motifs. *Nat Struct Mol Biol,* 19**,** 837-44.

Betous, R., Couch, F. B., Mason, A. C., Eichman, B. F., Manosas, M. & Cortez, D. 2013. Substrate-Selective Repair And Restart Of Replication Forks By DNA Translocases. *Cell Rep,* 3**,** 1958-69.

Bewersdorf, J., Bennett, B. T. & Knight, K. L. 2006. H2ax Chromatin Structures And Their Response To DNA Damage Revealed By 4pi Microscopy. *Proc Natl Acad Sci U S A,* 103**,** 18137-42.

Bhowmick, R., Minocherhomji, S. & Hickson, I. D. 2016. Rad52 Facilitates Mitotic DNA Synthesis Following Replication Stress. *Mol Cell,* 64**,** 1117-1126.

Bizard, A. H. & Hickson, I. D. 2014. The Dissolution Of Double Holliday Junctions. *Cold Spring Harb Perspect Biol,* 6**,** A016477.

Blackford, A. N. & Jackson, S. P. 2017. ATM, ATR, And DNA-PK: The Trinity At The Heart Of The DNA Damage Response. *Mol Cell,* 66**,** 801-817.

Blow, J. J. & Dutta, A. 2005. Preventing Re-Replication Of Chromosomal DNA. *Nat Rev Mol Cell Biol,* 6**,** 476-86.

Blow, J. J. & Ge, X. Q. 2008. Replication Forks, Chromatin Loops And Dormant Replication Origins. *Genome Biology,* 9**,** 244.

Blow, J. J. & Ge, X. Q. 2009. A Model For DNA Replication Showing How Dormant Origins Safeguard Against Replication Fork Failure. *Embo Reports,* 10**,** 406-12.

Blow, J. J., Ge, X. Q. & Jackson, D. A. 2011. How Dormant Origins Promote Complete Genome Replication. *Trends In Biochemical Sciences,* 36**,** 405-14.

Blow, J. J. & Gillespie, P. J. 2008. Replication Licensing And Cancer--A Fatal Entanglement? *Nat Rev Cancer,* 8**,** 799-806.

Blow, J. J. & Hodgson, B. 2002. Replication Licensing--Defining The Proliferative State? *Trends Cell Biol,* 12**,** 72-8.

Blow, J. J. & Laskey, R. A. 1988. A Role For The Nuclear Envelope In Controlling DNA Replication Within The Cell Cycle. *Nature,* 332**,** 546-8.

Bochman, M. L. & Schwacha, A. 2009. The Mcm Complex: Unwinding The Mechanism Of A Replicative Helicase. *Microbiol Mol Biol Rev,* 73**,** 652-83.

Boddy, M. N., Gaillard, P. H., Mcdonald, W. H., Shanahan, P., Yates, J. R., 3rd & Russell, P. 2001. Mus81-Eme1 Are Essential Components Of A Holliday Junction Resolvase. *Cell,* 107**,** 537-48.

Bork, P., Hofmann, K., Bucher, P., Neuwald, A. F., Altschul, S. F. & Koonin, E. V. 1997. A Superfamily Of Conserved Domains In DNA Damage-Responsive Cell Cycle Checkpoint Proteins. *Faseb J,* 11**,** 68-76.

Bothmer, A., Robbiani, D. F., Feldhahn, N., Gazumyan, A., Nussenzweig, A. & Nussenzweig, M. C. 2010. 53bp1 Regulates DNA Resection And The Choice Between Classical And Alternative End Joining During Class Switch Recombination. *J Exp Med,* 207**,** 855-65.

Bourseguin, J., Bonet, C., Renaud, E., Pandiani, C., Boncompagni, M., Giuliano, S., Pawlikowska, P., Karmous-Benailly, H., Ballotti, R., Rosselli, F. & Bertolotto, C. 2016. Fancd2 Functions As A Critical Factor Downstream Of Mitf To Maintain The Proliferation And Survival Of Melanoma Cells. *Sci Rep,* 6**,** 36539.

Bouwman, P., Aly, A., Escandell, J. M., Pieterse, M., Bartkova, J., Van Der Gulden, H., Hiddingh, S., Thanasoula, M., Kulkarni, A., Yang, Q., Haffty, B. G., Tommiska, J., Blomqvist, C., Drapkin, R., Adams, D. J., Nevanlinna, H., Bartek, J., Tarsounas, M., Ganesan, S. & Jonkers, J. 2010. 53bp1 Loss Rescues Brca1 Deficiency And Is Associated With Triple-Negative And Brca-Mutated Breast Cancers. *Nat Struct Mol Biol,* 17**,** 688-95.

Branzei, D. & Foiani, M. 2009. The Checkpoint Response To Replication Stress. *DNA Repair (Amst),* 8**,** 1038-46.

Brewer, B. J. & Fangman, W. L. 1993. Initiation At Closely Spaced Replication Origins In A Yeast Chromosome. *Science,* 262**,** 1728-31.

Brosh, R. M., Jr., Majumdar, A., Desai, S., Hickson, I. D., Bohr, V. A. & Seidman, M. M. 2001. Unwinding Of A DNA Triple Helix By The Werner And Bloom Syndrome Helicases. *J Biol Chem,* 276**,** 3024-30.

Brown, E. J. & Baltimore, D. 2000. ATR Disruption Leads To Chromosomal Fragmentation And Early Embryonic Lethality. *Genes Dev,* 14**,** 397-402.

Bunting, S. F., Callen, E., Wong, N., Chen, H. T., Polato, F., Gunn, A., Bothmer, A., Feldhahn, N., Fernandez-Capetillo, O., Cao, L., Xu, X., Deng, C. X., Finkel, T., Nussenzweig, M., Stark, J. M. & Nussenzweig, A. 2010. 53bp1 Inhibits Homologous Recombination In Brca1-Deficient Cells By Blocking Resection Of DNA Breaks. *Cell,* 141**,** 243-54.

Burkhart, R., Schulte, D., Hu, D., Musahl, C., Gohring, F. & Knippers, R. 1995. Interactions Of Human Nuclear Proteins P1mcm3 And P1cdc46. *Eur J Biochem,* 228**,** 431-8.

Callebaut, I. & Mornon, J. P. 1997. From Brca1 To Rap1: A Widespread Brct Module Closely Associated With DNA Repair. *Febs Lett,* 400**,** 25-30.

Callen, E., Di Virgilio, M., Kruhlak, M. J., Nieto-Soler, M., Wong, N., Chen, H. T., Faryabi, R. B., Polato, F., Santos, M., Starnes, L. M., Wesemann, D. R., Lee, J. E., Tubbs, A., Sleckman, B. P., Daniel, J. A., Ge, K., Alt, F. W., Fernandez-Capetillo, O., Nussenzweig, M. C. & Nussenzweig, A. 2013. 53bp1 Mediates Productive And Mutagenic DNA Repair Through Distinct Phosphoprotein Interactions. *Cell,* 153**,** 1266-80.

Cayrou, C., Ballester, B., Peiffer, I., Fenouil, R., Coulombe, P., Andrau, J. C., Van Helden, J. & Mechali, M. 2015. The Chromatin Environment Shapes DNA Replication Origin Organization And Defines Origin Classes. *Genome Res,* 25**,** 1873-85.

Chagin, V. O., Casas-Delucchi, C. S., Reinhart, M., Schermelleh, L., Markaki, Y., Maiser, A., Bolius, J. J., Bensimon, A., Fillies, M., Domaing, P., Rozanov, Y. M., Leonhardt, H. & Cardoso, M. C. 2016. 4d Visualization Of Replication Foci In Mammalian Cells Corresponding To Individual Replicons. *Nat Commun,* 7**,** 11231.

Chan, K. L. & Hickson, I. D. 2009. On The Origins Of Ultra-Fine Anaphase Bridges. *Cell Cycle,* 8**,** 3065-6.

Chan, K. L., North, P. S. & Hickson, I. D. 2007. Blm Is Required For Faithful Chromosome Segregation And Its Localization Defines A Class Of Ultrafine Anaphase Bridges. *Embo J,* 26**,** 3397-409.

Chan, K. L., Palmai-Pallag, T., Ying, S. & Hickson, I. D. 2009. Replication Stress Induces Sister-Chromatid Bridging At Fragile Site Loci In Mitosis. *Nat Cell Biol,* 11**,** 753-60.

Chapman, J. R., Barral, P., Vannier, J. B., Borel, V., Steger, M., Tomas-Loba, A., Sartori, A. A., Adams, I. R., Batista, F. D. & Boulton, S. J. 2013. Rif1 Is Essential For 53bp1-Dependent Nonhomologous End Joining And Suppression Of DNA Double-Strand Break Resection. *Mol Cell,* 49**,** 858-71.

Chapman, J. R., Sossick, A. J., Boulton, S. J. & Jackson, S. P. 2012. Brca1-Associated Exclusion Of 53bp1 From DNA Damage Sites Underlies Temporal Control Of DNA Repair. *J Cell Sci,* 125**,** 3529-34.

Chaudhury, I., Sareen, A., Raghunandan, M. & Sobeck, A. 2013. Fancd2 Regulates Blm Complex Functions Independently Of Fanci To Promote Replication Fork Recovery. *Nucleic Acids Res,* 41**,** 6444-59.

Chong, J. P., Thommes, P. & Blow, J. J. 1996. The Role Of Mcm/P1 Proteins In The Licensing Of DNA Replication. *Trends Biochem Sci,* 21**,** 102-6.

Ciccia, A., Constantinou, A. & West, S. C. 2003. Identification And Characterization Of The Human Mus81-Eme1 Endonuclease. *J Biol Chem,* 278**,** 25172-8.

Clarke, D. J., Johnson, R. T. & Downes, C. S. 1993. Topoisomerase Ii Inhibition Prevents Anaphase Chromatid Segregation In Mammalian Cells Independently Of The Generation Of DNA Strand Breaks. *J Cell Sci,* 105 (Pt 2)**,** 563-9.

Conti, C., Sacca, B., Herrick, J., Lalou, C., Pommier, Y. & Bensimon, A. 2007. Replication Fork Velocities At Adjacent Replication Origins Are Coordinately Modified During DNA Replication In Human Cells. *Mol Biol Cell,* 18**,** 3059-67.

Cortez, D., Guntuku, S., Qin, J. & Elledge, S. J. 2001. ATR And ATRIP: Partners In Checkpoint Signaling. *Science,* 294**,** 1713-6.

Costa, A., Renault, L., Swuec, P., Petojevic, T., Pesavento, J. J., Ilves, I., Maclellan-Gibson, K., Fleck, R. A., Botchan, M. R. & Berger, J. M. 2014. DNA Binding Polarity, Dimerization, And Atpase Ring Remodeling In The Cmg Helicase Of The Eukaryotic Replisome. *Elife,* 3**,** E03273.

Coster, G., Frigola, J., Beuron, F., Morris, E. P. & Diffley, J. F. 2014. Origin Licensing Requires Atp Binding And Hydrolysis By The Mcm Replicative Helicase. *Mol Cell,* 55**,** 666-77.

Crasta, K., Ganem, N. J., Dagher, R., Lantermann, A. B., Ivanova, E. V., Pan, Y., Nezi, L., Protopopov, A., Chowdhury, D. & Pellman, D. 2012. DNA Breaks And Chromosome Pulverization From Errors In Mitosis. *Nature,* 482**,** 53-8.

Daigaku, Y., Keszthelyi, A., Muller, C. A., Miyabe, I., Brooks, T., Retkute, R., Hubank, M., Nieduszyski, C. A. & Carr, A. M. 2015. A Global Profile Of Replicative Polymerase Usage. *Nat Struct Mol Biol,* 22**,** 192-198.

Davidson, I. F., Li, A. & Blow, J. J. 2006. Deregulated Replication Licensing Causes DNA Fragmentation Consistent With Head-To-Tail Fork Collision. *Mol Cell,* 24**,** 433-43.

Davies, S. L., North, P. S. & Hickson, I. D. 2007. Role For Blm In Replication-Fork Restart And Suppression Of Origin Firing After Replicative Stress. *Nat Struct Mol Biol,* 14**,** 677-9.

De Klein, A., Muijtjens, M., Van Os, R., Verhoeven, Y., Smit, B., Carr, A. M., Lehmann, A. R. & Hoeijmakers, J. H. 2000. Targeted Disruption Of The Cell-Cycle Checkpoint Gene ATR Leads To Early Embryonic Lethality In Mice. *Curr Biol,* 10**,** 479-82.

Deans, A. J. & West, S. C. 2009. Fancm Connects The Genome Instability Disorders Bloom's Syndrome And Fanconi Anemia. *Mol Cell,* 36**,** 943-53.

Deegan, T. D. & Diffley, J. F. 2016. Mcm: One Ring To Rule Them All. *Curr Opin Struct Biol,* 37**,** 145-51.

Deem, A., Keszthelyi, A., Blackgrove, T., Vayl, A., Coffey, B., Mathur, R., Chabes, A. & Malkova, A. 2011. Break-Induced Replication Is Highly Inaccurate. *Plos Biol,* 9**,** E1000594.

Dewar, J. M., Budzowska, M. & Walter, J. C. 2015. The Mechanism Of DNA Replication Termination In Vertebrates. *Nature,* 525**,** 345-50.

Diamant, N., Hendel, A., Vered, I., Carell, T., Reissner, T., De Wind, N., Geacinov, N. & Livneh, Z. 2012. DNA Damage Bypass Operates In The S And G2 Phases Of The Cell Cycle And Exhibits Differential Mutagenicity. *Nucleic Acids Res,* 40**,** 170-80.

Dimitrova, D. S. & Gilbert, D. M. 1999. The Spatial Position And Replication Timing Of Chromosomal Domains Are Both Established In Early G1 Phase. *Mol Cell,* 4**,** 983-93.

Dimitrova, D. S., Todorov, I. T., Melendy, T. & Gilbert, D. M. 1999. Mcm2, But Not Rpa, Is A Component Of The Mammalian Early G1-Phase Prereplication Complex. *J Cell Biol,* 146**,** 709-22.

Doe, C. L., Ahn, J. S., Dixon, J. & Whitby, M. C. 2002. Mus81-Eme1 And Rqh1 Involvement In Processing Stalled And Collapsed Replication Forks. *J Biol Chem,* 277**,** 32753-9.

Donnianni, R. A. & Symington, L. S. 2013. Break-Induced Replication Occurs By Conservative DNA Synthesis. *Proc Natl Acad Sci U S A,* 110**,** 13475-80.

Durkin, S. G. & Glover, T. W. 2007. Chromosome Fragile Sites. *Annu Rev Genet,* 41**,** 169-92.

Edwards, M. C., Tutter, A. V., Cvetic, C., Gilbert, C. H., Prokhorova, T. A. & Walter, J. C. 2002. Mcm2-7 Complexes Bind Chromatin In A Distributed Pattern Surrounding The Origin Recognition Complex In Xenopus Egg Extracts. *J Biol Chem,* 277**,** 33049-57.

Ekholm-Reed, S., Mendez, J., Tedesco, D., Zetterberg, A., Stillman, B. & Reed, S. I. 2004. Deregulation Of Cyclin E In Human Cells Interferes With Prereplication Complex Assembly. *J Cell Biol,* 165**,** 789-800.

Elledge, S. J. & Spottswood, M. R. 1991. A New Human P34 Protein Kinase, Cdk2, Identified By Complementation Of A Cdc28 Mutation In Saccharomyces Cerevisiae, Is A Homolog Of Xenopus Eg1. *Embo J,* 10**,** 2653-9.

Escribano-Diaz, C., Orthwein, A., Fradet-Turcotte, A., Xing, M., Young, J. T., Tkac, J., Cook, M. A., Rosebrock, A. P., Munro, M., Canny, M. D., Xu, D. & Durocher, D. 2013. A Cell Cycle-Dependent Regulatory Circuit Composed Of 53bp1-Rif1 And Brca1-Ctip Controls DNA Repair Pathway Choice. *Mol Cell,* 49**,** 872-83.

Evrin, C., Clarke, P., Zech, J., Lurz, R., Sun, J., Uhle, S., Li, H., Stillman, B. & Speck, C. 2009. A Double-Hexameric Mcm2-7 Complex Is Loaded Onto Origin DNA During Licensing Of Eukaryotic DNA Replication. *Proc Natl Acad Sci U S A,* 106**,** 20240-5.

Falbo, K. B. & Shen, X. 2006. Chromatin Remodeling In DNA Replication. *J Cell Biochem,* 97**,** 684-9.

Feng, D., Tu, Z., Wu, W. & Liang, C. 2003. Inhibiting The Expression Of DNA Replication-Initiation Proteins Induces Apoptosis In Human Cancer Cells. *Cancer Res,* 63**,** 7356-64.

Feng, L., Fong, K. W., Wang, J., Wang, W. & Chen, J. 2013. Rif1 Counteracts Brca1-Mediated End Resection During DNA Repair. *J Biol Chem,* 288**,** 11135-43.

Forsburg, S. L. 2004. Eukaryotic Mcm Proteins: Beyond Replication Initiation. *Microbiol Mol Biol Rev,* 68**,** 109-31.

Foulk, M. S., Urban, J. M., Casella, C. & Gerbi, S. A. 2015. Characterizing And Controlling Intrinsic Biases Of Lambda Exonuclease In Nascent Strand Sequencing Reveals Phasing Between Nucleosomes And G-Quadruplex Motifs Around A Subset Of Human Replication Origins. *Genome Res,* 25**,** 725-35.

Fragkos, M., Ganier, O., Coulombe, P. & Mechali, M. 2015. DNA Replication Origin Activation In Space And Time. *Nat Rev Mol Cell Biol,* 16**,** 360-74.

Francis, L. I., Randell, J. C., Takara, T. J., Uchima, L. & Bell, S. P. 2009. Incorporation Into The Prereplicative Complex Activates The Mcm2-7 Helicase For Cdc7-Dbf4 Phosphorylation. *Genes Dev,* 23**,** 643-54.

Frattini, A., Fabbri, M., Valli, R., De Paoli, E., Montalbano, G., Gribaldo, L., Pasquali, F. & Maserati, E. 2015. High Variability Of Genomic Instability And Gene Expression Profiling In Different Hela Clones. *Sci Rep,* 5**,** 15377.

Frigola, J., Remus, D., Mehanna, A. & Diffley, J. F. 2013. Atpase-Dependent Quality Control Of DNA Replication Origin Licensing. *Nature,* 495**,** 339-43.

Fu, Y. V., Yardimci, H., Long, D. T., Ho, T. V., Guainazzi, A., Bermudez, V. P., Hurwitz, J., Van Oijen, A., Scharer, O. D. & Walter, J. C. 2011. Selective Bypass Of A Lagging Strand Roadblock By The Eukaryotic Replicative DNA Helicase. *Cell,* 146**,** 931-41.

Galanos, P., Vougas, K., Walter, D., Polyzos, A., Maya-Mendoza, A., Haagensen, E. J., Kokkalis, A., Roumelioti, F. M., Gagos, S., Tzetis, M., Canovas, B., Igea, A., Ahuja, A. K., Zellweger, R., Havaki, S., Kanavakis, E., Kletsas, D., Roninson, I. B., Garbis, S. D., Lopes, M., Nebreda, A., Thanos, D., Blow, J. J., Townsend, P., Sorensen, C. S., Bartek, J. & Gorgoulis, V. G. 2016. Chronic P53-Independent P21 Expression Causes Genomic Instability By Deregulating Replication Licensing. *Nat Cell Biol,* 18**,** 777-89.

Gambus, A., Jones, R. C., Sanchez-Diaz, A., Kanemaki, M., Van Deursen, F., Edmondson, R. D. & Labib, K. 2006. Gins Maintains Association Of Cdc45 With Mcm In Replisome Progression Complexes At Eukaryotic DNA Replication Forks. *Nat Cell Biol,* 8**,** 358-66.

Gambus, A., Van Deursen, F., Polychronopoulos, D., Foltman, M., Jones, R. C., Edmondson, R. D., Calzada, A. & Labib, K. 2009. A Key Role For Ctf4 In Coupling The Mcm2-7 Helicase To DNA Polymerase Alpha Within The Eukaryotic Replisome. *Embo J,* 28**,** 2992-3004.

Garcia-Gomez, S., Reyes, A., Martinez-Jimenez, M. I., Chocron, E. S., Mouron, S., Terrados, G., Powell, C., Salido, E., Mendez, J., Holt, I. J. & Blanco, L. 2013. Primpol, An Archaic Primase/Polymerase Operating In Human Cells. *Mol Cell,* 52**,** 541-53.

Garcia-Higuera, I., Taniguchi, T., Ganesan, S., Meyn, M. S., Timmers, C., Hejna, J., Grompe, M. & D'andrea, A. D. 2001. Interaction Of The Fanconi Anemia Proteins And Brca1 In A Common Pathway. *Mol Cell,* 7**,** 249-62.

Gardner, N. J., Gillespie, P. J., Carrington, J. T., Shanks, E. J., Mcelroy, S. P., Haagensen, E. J., Frearson, J. A., Woodland, A. & Blow, J. J. 2017. The High-Affinity Interaction Between Orc And DNA That Is Required For Replication Licensing Is Inhibited By 2-Arylquinolin-4-Amines. *Cell Chem Biol,* 24**,** 981-992 E4.

Garner, E. & Smogorzewska, A. 2011. Ubiquitylation And The Fanconi Anemia Pathway. *Febs Lett,* 585**,** 2853-60.

Ge, X. Q. & Blow, J. J. 2009. The Licensing Checkpoint Opens Up. *Cell Cycle,* 8**,** 2320-2.

Ge, X. Q. & Blow, J. J. 2010. Chk1 Inhibits Replication Factory Activation But Allows Dormant Origin Firing In Existing Factories. *The Journal Of Cell Biology,* 191**,** 1285-97.

Ge, X. Q., Jackson, D. A. & Blow, J. J. 2007. Dormant Origins Licensed By Excess Mcm2-7 Are Required For Human Cells To Survive Replicative Stress. *Genes & Development,* 21**,** 3331-41.

Geng, Y., Yu, Q., Sicinska, E., Das, M., Schneider, J. E., Bhattacharya, S., Rideout, W. M., Bronson, R. T., Gardner, H. & Sicinski, P. 2003. Cyclin E Ablation In The Mouse. *Cell,* 114**,** 431-43.

Georgescu, R. E., Langston, L., Yao, N. Y., Yurieva, O., Zhang, D., Finkelstein, J., Agarwal, T. & O'donnell, M. E. 2014. Mechanism Of Asymmetric Polymerase Assembly At The Eukaryotic Replication Fork. *Nat Struct Mol Biol,* 21**,** 664-70.

German, J. 1993. Bloom Syndrome: A Mendelian Prototype Of Somatic Mutational Disease. *Medicine (Baltimore),* 72**,** 393-406.

Gilbert, D. M. 2001. Making Sense Of Eukaryotic DNA Replication Origins. *Science,* 294**,** 96-100.

Gillespie, P. J., Li, A. & Blow, J. J. 2001. Reconstitution Of Licensed Replication Origins On Xenopus Sperm Nuclei Using Purified Proteins. *Bmc Biochem,* 2**,** 15.

Giunta, S., Belotserkovskaya, R. & Jackson, S. P. 2010. DNA Damage Signaling In Response To Double-Strand Breaks During Mitosis. *J Cell Biol,* 190**,** 197-207.

Glover, T. W., Berger, C., Coyle, J. & Echo, B. 1984. DNA Polymerase Alpha Inhibition By Aphidicolin Induces Gaps And Breaks At Common Fragile Sites In Human Chromosomes. *Hum Genet,* 67**,** 136-42.

Green, B. M. & Li, J. J. 2005. Loss Of Rereplication Control In Saccharomyces Cerevisiae Results In Extensive DNA Damage. *Mol Biol Cell,* 16**,** 421-32.

Guillou, E., Ibarra, A., Coulon, V., Casado-Vela, J., Rico, D., Casal, I., Schwob, E., Losada, A. & Mendez, J. 2010. Cohesin Organizes Chromatin Loops At DNA Replication Factories. *Genes Dev,* 24**,** 2812-22.

Hanada, K., Budzowska, M., Davies, S. L., Van Drunen, E., Onizawa, H., Beverloo, H. B., Maas, A., Essers, J., Hickson, I. D. & Kanaar, R. 2007. The Structure-Specific Endonuclease Mus81 Contributes To Replication Restart By Generating Double-Strand DNA Breaks. *Nat Struct Mol Biol,* 14**,** 1096-104.

Hanna, J. S., Kroll, E. S., Lundblad, V. & Spencer, F. A. 2001. Saccharomyces Cerevisiae Ctf18 And Ctf4 Are Required For Sister Chromatid Cohesion. *Mol Cell Biol,* 21**,** 3144-58.

Hardy, C. F., Dryga, O., Seematter, S., Pahl, P. M. & Sclafani, R. A. 1997. Mcm5/Cdc46-Bob1 Bypasses The Requirement For The S Phase Activator Cdc7p. *Proc Natl Acad Sci U S A,* 94**,** 3151-5.

Harrigan, J. A., Belotserkovskaya, R., Coates, J., Dimitrova, D. S., Polo, S. E., Bradshaw, C. R., Fraser, P. & Jackson, S. P. 2011. Replication Stress Induces 53bp1-Containing Opt Domains In G1 Cells. *J Cell Biol,* 193**,** 97-108.

Hemphill, A. W., Akkari, Y., Newell, A. H., Schultz, R. A., Grompe, M., North, P. S., Hickson, I. D., Jakobs, P. M., Rennie, S., Pauw, D., Hejna, J., Olson, S. B. & Moses, R. E. 2009. Topo Iiialpha And Blm Act Within The Fanconi Anemia Pathway In Response To DNA-Crosslinking Agents. *Cytogenet Genome Res,* 125**,** 165-75.

Heyer, W. D., Ehmsen, K. T. & Liu, J. 2010. Regulation Of Homologous Recombination In Eukaryotes. *Annu Rev Genet,* 44**,** 113-39.

Hofmann, J. F. & Beach, D. 1994. Cdt1 Is An Essential Target Of The Cdc10/Sct1 Transcription Factor: Requirement For DNA Replication And Inhibition Of Mitosis. *Embo J,* 13**,** 425-34.

Honda, R. & Yasuda, H. 1999. Association Of P19(Arf) With Mdm2 Inhibits Ubiquitin Ligase Activity Of Mdm2 For Tumor Suppressor P53. *Embo J,* 18**,** 22-7.

Hook, S. S., Lin, J. J. & Dutta, A. 2007. Mechanisms To Control Rereplication And Implications For Cancer. *Curr Opin Cell Biol,* 19**,** 663-71.

Hooks, K. B., Delneri, D. & Griffiths-Jones, S. 2014. Intron Evolution In Saccharomycetaceae. *Genome Biol Evol,* 6**,** 2543-56.

Hyrien, O., Marheineke, K. & Goldar, A. 2003. Paradoxes Of Eukaryotic DNA Replication: Mcm Proteins And The Random Completion Problem. *Bioessays,* 25**,** 116-25.

Ibarra, A., Schwob, E. & Mendez, J. 2008. Excess Mcm Proteins Protect Human Cells From Replicative Stress By Licensing Backup Origins Of Replication. *Proceedings Of The National Academy Of Sciences Of The United States Of America,* 105**,** 8956-61.

Ishimi, Y., Okayasu, I., Kato, C., Kwon, H. J., Kimura, H., Yamada, K. & Song, S. Y. 2003. Enhanced Expression Of Mcm Proteins In Cancer Cells Derived From Uterine Cervix. *Eur J Biochem,* 270**,** 1089-101.

Iwabuchi, K., Bartel, P. L., Li, B., Marraccino, R. & Fields, S. 1994. Two Cellular Proteins That Bind To Wild-Type But Not Mutant P53. *Proc Natl Acad Sci U S A,* 91**,** 6098-102.

Jackson, D. A. & Pombo, A. 1998. Replicon Clusters Are Stable Units Of Chromosome Structure: Evidence That Nuclear Organization Contributes To The Efficient Activation And Propagation Of S Phase In Human Cells. *J Cell Biol,* 140**,** 1285-95.

Janssen, A. & Medema, R. H. 2013. Genetic Instability: Tipping The Balance. *Oncogene,* 32**,** 4459-70.

Jares, P. & Blow, J. J. 2000. Xenopus Cdc7 Function Is Dependent On Licensing But Not On Xorc, Xcdc6, Or Cdk Activity And Is Required For Xcdc45 Loading. *Genes Dev,* 14**,** 1528-40.

Jin, J., Arias, E. E., Chen, J., Harper, J. W. & Walter, J. C. 2006. A Family Of Diverse Cul4-Ddb1-Interacting Proteins Includes Cdt2, Which Is Required For S Phase Destruction Of The Replication Factor Cdt1. *Mol Cell,* 23**,** 709-21.

Jullien, D., Vagnarelli, P., Earnshaw, W. C. & Adachi, Y. 2002. Kinetochore Localisation Of The DNA Damage Response Component 53bp1 During Mitosis. *J Cell Sci,* 115**,** 71-9.

Kang, Y. H., Galal, W. C., Farina, A., Tappin, I. & Hurwitz, J. 2012. Properties Of The Human Cdc45/Mcm2-7/Gins Helicase Complex And Its Action With DNA Polymerase Epsilon In Rolling Circle DNA Synthesis. *Proc Natl Acad Sci U S A,* 109**,** 6042-7.

Karakaidos, P., Taraviras, S., Vassiliou, L. V., Zacharatos, P., Kastrinakis, N. G., Kougiou, D., Kouloukoussa, M., Nishitani, H., Papavassiliou, A. G., Lygerou, Z. & Gorgoulis, V. G. 2004. Overexpression Of The Replication Licensing Regulators Hcdt1 And Hcdc6 Characterizes A Subset Of Non-Small-Cell Lung Carcinomas: Synergistic Effect With Mutant P53 On Tumor Growth And Chromosomal Instability--Evidence Of E2f-1 Transcriptional Control Over Hcdt1. *Am J Pathol,* 165**,** 1351-65.

Karow, J. K., Chakraverty, R. K. & Hickson, I. D. 1997. The Bloom's Syndrome Gene Product Is A 3'-5' DNA Helicase. *J Biol Chem,* 272**,** 30611-4.

Karow, J. K., Constantinou, A., Li, J. L., West, S. C. & Hickson, I. D. 2000. The Bloom's Syndrome Gene Product Promotes Branch Migration Of Holliday Junctions. *Proc Natl Acad Sci U S A,* 97**,** 6504-8.

Kee, Y. & D'andrea, A. D. 2010. Expanded Roles Of The Fanconi Anemia Pathway In Preserving Genomic Stability. *Genes Dev,* 24**,** 1680-94.

Kee, Y. & D'andrea, A. D. 2012. Molecular Pathogenesis And Clinical Management Of Fanconi Anemia. *J Clin Invest,* 122**,** 3799-806.

Kozar, K., Ciemerych, M. A., Rebel, V. I., Shigematsu, H., Zagozdzon, A., Sicinska, E., Geng, Y., Yu, Q., Bhattacharya, S., Bronson, R. T., Akashi, K. & Sicinski, P. 2004. Mouse Development And Cell Proliferation In The Absence Of D-Cyclins. *Cell,* 118**,** 477-91.

Krude, T., Musahl, C., Laskey, R. A. & Knippers, R. 1996. Human Replication Proteins Hcdc21, Hcdc46 And P1mcm3 Bind Chromatin Uniformly Before S-Phase And Are Displaced Locally During DNA Replication. *J Cell Sci,* 109 (Pt 2)**,** 309-18.

Kunnev, D., Freeland, A., Qin, M., Leach, R. W., Wang, J., Shenoy, R. M. & Pruitt, S. C. 2015. Effect Of Minichromosome Maintenance Protein 2 Deficiency On The Locations Of DNA Replication Origins. *Genome Res,* 25**,** 558-69.

Kurat, C. F., Yeeles, J. T., Patel, H., Early, A. & Diffley, J. F. 2017. Chromatin Controls DNA Replication Origin Selection, Lagging-Strand Synthesis, And Replication Fork Rates. *Mol Cell,* 65**,** 117-130.

Labib, K. & Diffley, J. F. 2001. Is The Mcm2-7 Complex The Eukaryotic DNA Replication Fork Helicase? *Curr Opin Genet Dev,* 11**,** 64-70.

Labib, K., Tercero, J. A. & Diffley, J. F. 2000. Uninterrupted Mcm2-7 Function Required For DNA Replication Fork Progression. *Science,* 288**,** 1643-7.

Lachaud, C., Moreno, A., Marchesi, F., Toth, R., Blow, J. J. & Rouse, J. 2016. Ubiquitinated Fancd2 Recruits Fan1 To Stalled Replication Forks To Prevent Genome Instability. *Science,* 351**,** 846-9.

Lebofsky, R., Heilig, R., Sonnleitner, M., Weissenbach, J. & Bensimon, A. 2006. DNA Replication Origin Interference Increases The Spacing Between Initiation Events In Human Cells. *Mol Biol Cell,* 17**,** 5337-45.

Lee, C., Hong, B., Choi, J. M., Kim, Y., Watanabe, S., Ishimi, Y., Enomoto, T., Tada, S., Kim, Y. & Cho, Y. 2004. Structural Basis For Inhibition Of The Replication Licensing Factor Cdt1 By Geminin. *Nature,* 430**,** 913-7.

Lei, M., Kawasaki, Y. & Tye, B. K. 1996. Physical Interactions Among Mcm Proteins And Effects Of Mcm Dosage On DNA Replication In Saccharomyces Cerevisiae. *Mol Cell Biol,* 16**,** 5081-90.

Leonhardt, H., Rahn, H. P., Weinzierl, P., Sporbert, A., Cremer, T., Zink, D. & Cardoso, M. C. 2000. Dynamics Of DNA Replication Factories In Living Cells. *J Cell Biol,* 149**,** 271-80.

Letessier, A., Millot, G. A., Koundrioukoff, S., Lachages, A. M., Vogt, N., Hansen, R. S., Malfoy, B., Brison, O. & Debatisse, M. 2011. Cell-Type-Specific Replication Initiation Programs Set Fragility Of The Fra3b Fragile Site. *Nature,* 470**,** 120-3.

Li, A. & Blow, J. J. 2005. Cdt1 Downregulation By Proteolysis And Geminin Inhibition Prevents DNA Re-Replication In Xenopus. *Embo J,* 24**,** 395-404.

Li, J. J. & Herskowitz, I. 1993. Isolation Of Orc6, A Component Of The Yeast Origin Recognition Complex By A One-Hybrid System. *Science,* 262**,** 1870-4.

Li, N., Zhai, Y., Zhang, Y., Li, W., Yang, M., Lei, J., Tye, B. K. & Gao, N. 2015. Structure Of The Eukaryotic Mcm Complex At 3.8 A. *Nature,* 524**,** 186-91.

Lieber, M. R. 2010. The Mechanism Of Double-Strand DNA Break Repair By The Nonhomologous DNA End-Joining Pathway. *Annu Rev Biochem,* 79**,** 181-211.

Liku, M. E., Nguyen, V. Q., Rosales, A. W., Irie, K. & Li, J. J. 2005. Cdk Phosphorylation Of A Novel Nls-Nes Module Distributed Between Two Subunits Of The Mcm2-7 Complex Prevents Chromosomal Rereplication. *Mol Biol Cell,* 16**,** 5026-39.

Liontos, M., Koutsami, M., Sideridou, M., Evangelou, K., Kletsas, D., Levy, B., Kotsinas, A., Nahum, O., Zoumpourlis, V., Kouloukoussa, M., Lygerou, Z., Taraviras, S., Kittas, C., Bartkova, J., Papavassiliou, A. G., Bartek, J., Halazonetis, T. D. & Gorgoulis, V. G. 2007. Deregulated Overexpression Of Hcdt1 And Hcdc6 Promotes Malignant Behavior. *Cancer Res,* 67**,** 10899-909.

Liu, Y., Nielsen, C. F., Yao, Q. & Hickson, I. D. 2014. The Origins And Processing Of Ultra Fine Anaphase DNA Bridges. *Curr Opin Genet Dev,* 26**,** 1-5.

Lob, D., Lengert, N., Chagin, V. O., Reinhart, M., Casas-Delucchi, C. S., Cardoso, M. C. & Drossel, B. 2016. 3d Replicon Distributions Arise From Stochastic Initiation And Domino-Like DNA Replication Progression. *Nat Commun,* 7**,** 11207.

Loeb, K. R., Kostner, H., Firpo, E., Norwood, T., K, D. T., Clurman, B. E. & Roberts, J. M. 2005. A Mouse Model For Cyclin E-Dependent Genetic Instability And Tumorigenesis. *Cancer Cell,* 8**,** 35-47.

Lopes, M., Foiani, M. & Sogo, J. M. 2006. Multiple Mechanisms Control Chromosome Integrity After Replication Fork Uncoupling And Restart At Irreparable Uv Lesions. *Mol Cell,* 21**,** 15-27.

Lovejoy, C. A., Lock, K., Yenamandra, A. & Cortez, D. 2006. Ddb1 Maintains Genome Integrity Through Regulation Of Cdt1. *Mol Cell Biol,* 26**,** 7977-90.

Lukas, C., Savic, V., Bekker-Jensen, S., Doil, C., Neumann, B., Pedersen, R. S., Grofte, M., Chan, K. L., Hickson, I. D., Bartek, J. & Lukas, J. 2011. 53bp1 Nuclear Bodies Form Around DNA Lesions Generated By Mitotic Transmission Of Chromosomes Under Replication Stress. *Nat Cell Biol,* 13**,** 243-53.

Lydeard, J. R., Jain, S., Yamaguchi, M. & Haber, J. E. 2007. Break-Induced Replication And Telomerase-Independent Telomere Maintenance Require Pol32. *Nature,* 448**,** 820-3.

Lydeard, J. R., Lipkin-Moore, Z., Sheu, Y. J., Stillman, B., Burgers, P. M. & Haber, J. E. 2010. Break-Induced Replication Requires All Essential DNA Replication Factors Except Those Specific For Pre-Rc Assembly. *Genes Dev,* 24**,** 1133-44.

Mackay, C., Declais, A. C., Lundin, C., Agostinho, A., Deans, A. J., Macartney, T. J., Hofmann, K., Gartner, A., West, S. C., Helleday, T., Lilley, D. M. & Rouse, J. 2010. Identification Of Kiaa1018/Fan1, A DNA Repair Nuclease Recruited To DNA Damage By Monoubiquitinated Fancd2. *Cell,* 142**,** 65-76.

Macville, M., Schrock, E., Padilla-Nash, H., Keck, C., Ghadimi, B. M., Zimonjic, D., Popescu, N. & Ried, T. 1999. Comprehensive And Definitive Molecular Cytogenetic Characterization Of Hela Cells By Spectral Karyotyping. *Cancer Res,* 59**,** 141-50.

Madine, M. A., Khoo, C. Y., Mills, A. D., Musahl, C. & Laskey, R. A. 1995. The Nuclear Envelope Prevents Reinitiation Of Replication By Regulating The Binding Of Mcm3 To Chromatin In Xenopus Egg Extracts. *Curr Biol,* 5**,** 1270-9.

Mahbubani, H. M., Chong, J. P., Chevalier, S., Thommes, P. & Blow, J. J. 1997. Cell Cycle Regulation Of The Replication Licensing System: Involvement Of A Cdk-Dependent Inhibitor. *J Cell Biol,* 136**,** 125-35.

Maine, G. T., Sinha, P. & Tye, B. K. 1984. Mutants Of S. Cerevisiae Defective In The Maintenance Of Minichromosomes. *Genetics,* 106**,** 365-85.

Malkova, A., Naylor, M. L., Yamaguchi, M., Ira, G. & Haber, J. E. 2005. Rad51-Dependent Break-Induced Replication Differs In Kinetics And Checkpoint Responses From Rad51-Mediated Gene Conversion. *Mol Cell Biol,* 25**,** 933-44.

Malumbres, M. 2014. Cyclin-Dependent Kinases. *Genome Biol,* 15**,** 122.

Malumbres, M. & Barbacid, M. 2005. Mammalian Cyclin-Dependent Kinases. *Trends Biochem Sci,* 30**,** 630-41.

Mankouri, H. W. & Hickson, I. D. 2007. The Recq Helicase-Topoisomerase Iii-Rmi1 Complex: A DNA Structure-Specific 'Dissolvasome'? *Trends Biochem Sci,* 32**,** 538-46.

Mankouri, H. W., Huttner, D. & Hickson, I. D. 2013. How Unfinished Business From S-Phase Affects Mitosis And Beyond. *Embo J,* 32**,** 2661-71.

Mao, Z., Bozzella, M., Seluanov, A. & Gorbunova, V. 2008a. Comparison Of Nonhomologous End Joining And Homologous Recombination In Human Cells. *DNA Repair (Amst),* 7**,** 1765-71.

Mao, Z., Bozzella, M., Seluanov, A. & Gorbunova, V. 2008b. DNA Repair By Nonhomologous End Joining And Homologous Recombination During Cell Cycle In Human Cells. *Cell Cycle,* 7**,** 2902-6.

Marechal, A. & Zou, L. 2013. DNA Damage Sensing By The ATM And ATR Kinases. *Cold Spring Harb Perspect Biol,* 5.

Marheineke, K. & Hyrien, O. 2004. Control Of Replication Origin Density And Firing Time In Xenopus Egg Extracts: Role Of A Caffeine-Sensitive, ATR-Dependent Checkpoint. *J Biol Chem,* 279**,** 28071-81.

Maric, M., Maculins, T., De Piccoli, G. & Labib, K. 2014. Cdc48 And A Ubiquitin Ligase Drive Disassembly Of The Cmg Helicase At The End Of DNA Replication. *Science,* 346**,** 1253596.

Martin, M. M., Ryan, M., Kim, R., Zakas, A. L., Fu, H., Lin, C. M., Reinhold, W. C., Davis, S. R., Bilke, S., Liu, H., Doroshow, J. H., Reimers, M. A., Valenzuela, M. S., Pommier, Y., Meltzer, P. S. & Aladjem, M. I. 2011. Genome-Wide Depletion Of Replication Initiation Events In Highly Transcribed Regions. *Genome Res,* 21**,** 1822-32.

Masai, H., Matsumoto, S., You, Z., Yoshizawa-Sugata, N. & Oda, M. 2010. Eukaryotic Chromosome DNA Replication: Where, When, And How? *Annu Rev Biochem,* 79**,** 89-130.

Mceachern, M. J. & Haber, J. E. 2006. Break-Induced Replication And Recombinational Telomere Elongation In Yeast. *Annu Rev Biochem,* 75**,** 111-35.

Meetei, A. R., Sechi, S., Wallisch, M., Yang, D., Young, M. K., Joenje, H., Hoatlin, M. E. & Wang, W. 2003. A Multiprotein Nuclear Complex Connects Fanconi Anemia And Bloom Syndrome. *Mol Cell Biol,* 23**,** 3417-26.

Melixetian, M., Ballabeni, A., Masiero, L., Gasparini, P., Zamponi, R., Bartek, J., Lukas, J. & Helin, K. 2004. Loss Of Geminin Induces Rereplication In The Presence Of Functional P53. *J Cell Biol,* 165**,** 473-82.

Mendez, J., Zou-Yang, X. H., Kim, S. Y., Hidaka, M., Tansey, W. P. & Stillman, B. 2002. Human Origin Recognition Complex Large Subunit Is Degraded By Ubiquitin-Mediated Proteolysis After Initiation Of DNA Replication. *Mol Cell,* 9**,** 481-91.

Meselson, M. & Stahl, F. W. 1958. The Replication Of DNA In Escherichia Coli. *Proc Natl Acad Sci U S A,* 44**,** 671-82.

Mesner, L. D., Crawford, E. L. & Hamlin, J. L. 2006. Isolating Apparently Pure Libraries Of Replication Origins From Complex Genomes. *Mol Cell,* 21**,** 719-26.

Mesner, L. D., Valsakumar, V., Cieslik, M., Pickin, R., Hamlin, J. L. & Bekiranov, S. 2013. Bubble-Seq Analysis Of The Human Genome Reveals Distinct Chromatin-Mediated Mechanisms For Regulating Early- And Late-Firing Origins. *Genome Res,* 23**,** 1774-88.

Minocherhomji, S., Ying, S., Bjerregaard, V. A., Bursomanno, S., Aleliunaite, A., Wu, W., Mankouri, H. W., Shen, H., Liu, Y. & Hickson, I. D. 2015. Replication Stress Activates DNA Repair Synthesis In Mitosis. *Nature,* 528**,** 286-90.

Mohaghegh, P., Karow, J. K., Brosh, R. M., Jr., Bohr, V. A. & Hickson, I. D. 2001. The Bloom's And Werner's Syndrome Proteins Are DNA Structure-Specific Helicases. *Nucleic Acids Res,* 29**,** 2843-9.

Moldovan, G. L. & D'andrea, A. D. 2009. How The Fanconi Anemia Pathway Guards The Genome. *Annu Rev Genet,* 43**,** 223-49.

Montagnoli, A., Tenca, P., Sola, F., Carpani, D., Brotherton, D., Albanese, C. & Santocanale, C. 2004. Cdc7 Inhibition Reveals A P53-Dependent Replication Checkpoint That Is Defective In Cancer Cells. *Cancer Res,* 64**,** 7110-6.

Moore, J. D., Kirk, J. A. & Hunt, T. 2003. Unmasking The S-Phase-Promoting Potential Of Cyclin B1. *Science,* 300**,** 987-90.

Morales, J. C., Xia, Z., Lu, T., Aldrich, M. B., Wang, B., Rosales, C., Kellems, R. E., Hittelman, W. N., Elledge, S. J. & Carpenter, P. B. 2003. Role For The Brca1 C-Terminal Repeats (Brct) Protein 53bp1 In Maintaining Genomic Stability. *J Biol Chem,* 278**,** 14971-7.

Moreno, A., Carrington, J. T., Albergante, L., Al Mamun, M., Haagensen, E. J., Komseli, E. S., Gorgoulis, V. G., Newman, T. J. & Blow, J. J. 2016. Unreplicated DNA Remaining From Unperturbed S Phases Passes Through Mitosis For Resolution In Daughter Cells. *Proceedings Of The National Academy Of Sciences Of The United States Of America,* 113**,** E5757-64.

Moreno, S. P., Bailey, R., Campion, N., Herron, S. & Gambus, A. 2014. Polyubiquitylation Drives Replisome Disassembly At The Termination Of DNA Replication. *Science,* 346**,** 477-81.

Morohashi, H., Maculins, T. & Labib, K. 2009. The Amino-Terminal Tpr Domain Of Dia2 Tethers Scf(Dia2) To The Replisome Progression Complex. *Curr Biol,* 19**,** 1943-9.

Mouron, S., Rodriguez-Acebes, S., Martinez-Jimenez, M. I., Garcia-Gomez, S., Chocron, S., Blanco, L. & Mendez, J. 2013. Repriming Of DNA Synthesis At Stalled Replication Forks By Human Primpol. *Nat Struct Mol Biol,* 20**,** 1383-9.

Moyer, S. E., Lewis, P. W. & Botchan, M. R. 2006. Isolation Of The Cdc45/Mcm2-7/Gins (Cmg) Complex, A Candidate For The Eukaryotic DNA Replication Fork Helicase. *Proc Natl Acad Sci U S A,* 103**,** 10236-41.

Munoz, S., Bua, S., Rodriguez-Acebes, S., Megias, D., Ortega, S., De Martino, A. & Mendez, J. 2017. In Vivo DNA Re-Replication Elicits Lethal Tissue Dysplasias. *Cell Rep,* 19**,** 928-938.

Murga, M., Lecona, E., Kamileri, I., Diaz, M., Lugli, N., Sotiriou, S. K., Anton, M. E., Mendez, J., Halazonetis, T. D. & Fernandez-Capetillo, O. 2016. Pold3 Is Haploinsufficient For DNA Replication In Mice. *Mol Cell,* 63**,** 877-83.

Murray, A. W. 2004. Recycling The Cell Cycle: Cyclins Revisited. *Cell,* 116**,** 221-34.

Musahl, C., Holthoff, H. P., Lesch, R. & Knippers, R. 1998. Stability Of The Replicative Mcm3 Protein In Proliferating And Differentiating Human Cells. *Exp Cell Res,* 241**,** 260-4.

Naim, V. & Rosselli, F. 2009. The Fanc Pathway And Blm Collaborate During Mitosis To Prevent Micro-Nucleation And Chromosome Abnormalities. *Nat Cell Biol,* 11**,** 761-8.

Natale, F., Rapp, A., Yu, W., Maiser, A., Harz, H., Scholl, A., Grulich, S., Anton, T., Horl, D., Chen, W., Durante, M., Taucher-Scholz, G., Leonhardt, H. & Cardoso, M. C. 2017. Identification Of The Elementary Structural Units Of The DNA Damage Response. *Nat Commun,* 8**,** 15760.

Natsume, T., Kiyomitsu, T., Saga, Y. & Kanemaki, M. T. 2016. Rapid Protein Depletion In Human Cells By Auxin-Inducible Degron Tagging With Short Homology Donors. *Cell Rep,* 15**,** 210-8.

Natsume, T., Nishimura, K., Minocherhomji, S., Bhowmick, R., Hickson, I. D. & Kanemaki, M. T. 2017. Acute Inactivation Of The Replicative Helicase In Human Cells Triggers Mcm8-9-Dependent DNA Synthesis. *Genes Dev,* 31**,** 816-829.

Neelsen, K. J. & Lopes, M. 2015. Replication Fork Reversal In Eukaryotes: From Dead End To Dynamic Response. *Nat Rev Mol Cell Biol,* 16**,** 207-20.

Nelson, G., Buhmann, M. & Von Zglinicki, T. 2009. DNA Damage Foci In Mitosis Are Devoid Of 53bp1. *Cell Cycle,* 8**,** 3379-83.

Nevis, K. R., Cordeiro-Stone, M. & Cook, J. G. 2009. Origin Licensing And P53 Status Regulate Cdk2 Activity During G(1). *Cell Cycle,* 8**,** 1952-63.

Newman, T. J., Mamun, M. A., Nieduszynski, C. A. & Blow, J. J. 2013. Replisome Stall Events Have Shaped The Distribution Of Replication Origins In The Genomes Of Yeasts. *Nucleic Acids Research,* 41**,** 9705-18.

Nguyen, V. Q., Co, C. & Li, J. J. 2001. Cyclin-Dependent Kinases Prevent DNA Re-Replication Through Multiple Mechanisms. *Nature,* 411**,** 1068-73.

Nichols, W. W., Murphy, D. G., Cristofalo, V. J., Toji, L. H., Greene, A. E. & Dwight, S. A. 1977. Characterization Of A New Human Diploid Cell Strain, Imr-90. *Science,* 196**,** 60-3.

Nicholson, J. M. & Cimini, D. 2013. Cancer Karyotypes: Survival Of The Fittest. *Front Oncol,* 3**,** 148.

Nieduszynski, C. A., Hiraga, S., Ak, P., Benham, C. J. & Donaldson, A. D. 2007. Oridb: A DNA Replication Origin Database. *Nucleic Acids Res,* 35**,** D40-6.

Nijman, S. M., Huang, T. T., Dirac, A. M., Brummelkamp, T. R., Kerkhoven, R. M., D'andrea, A. D. & Bernards, R. 2005. The Deubiquitinating Enzyme Usp1 Regulates The Fanconi Anemia Pathway. *Mol Cell,* 17**,** 331-9.

Nishitani, H., Lygerou, Z., Nishimoto, T. & Nurse, P. 2000. The Cdt1 Protein Is Required To License DNA For Replication In Fission Yeast. *Nature,* 404**,** 625-8.

Nishitani, H., Taraviras, S., Lygerou, Z. & Nishimoto, T. 2001. The Human Licensing Factor For DNA Replication Cdt1 Accumulates In G1 And Is Destabilized After Initiation Of S-Phase. *J Biol Chem,* 276**,** 44905-11.

O'keefe, R. T., Henderson, S. C. & Spector, D. L. 1992. Dynamic Organization Of DNA Replication In Mammalian Cell Nuclei: Spatially And Temporally Defined Replication Of Chromosome-Specific Alpha-Satellite DNA Sequences. *J Cell Biol,* 116**,** 1095-110.

Oehlmann, M., Score, A. J. & Blow, J. J. 2004. The Role Of Cdc6 In Ensuring Complete Genome Licensing And S Phase Checkpoint Activation. *J Cell Biol,* 165**,** 181-90.

On, K. F., Beuron, F., Frith, D., Snijders, A. P., Morris, E. P. & Diffley, J. F. 2014. Prereplicative Complexes Assembled In Vitro Support Origin-Dependent And Independent DNA Replication. *Embo J,* 33**,** 605-20.

Palumbo, E., Matricardi, L., Tosoni, E., Bensimon, A. & Russo, A. 2010. Replication Dynamics At Common Fragile Site Fra6e. *Chromosoma,* 119**,** 575-87.

Panier, S. & Boulton, S. J. 2014. Double-Strand Break Repair: 53bp1 Comes Into Focus. *Nat Rev Mol Cell Biol,* 15**,** 7-18.

Paris, J., Le Guellec, R., Couturier, A., Le Guellec, K., Omilli, F., Camonis, J., Macneill, S. & Philippe, M. 1991. Cloning By Differential Screening Of A Xenopus Cdna Coding For A Protein Highly Homologous To Cdc2. *Proc Natl Acad Sci U S A,* 88**,** 1039-43.

Pepe, A. & West, S. C. 2014. Mus81-Eme2 Promotes Replication Fork Restart. *Cell Rep,* 7**,** 1048-55.

Petermann, E., Maya-Mendoza, A., Zachos, G., Gillespie, D. A., Jackson, D. A. & Caldecott, K. W. 2006. Chk1 Requirement For High Global Rates Of Replication Fork Progression During Normal Vertebrate S Phase. *Mol Cell Biol,* 26**,** 3319-26.

Petermann, E., Orta, M. L., Issaeva, N., Schultz, N. & Helleday, T. 2010. Hydroxyurea-Stalled Replication Forks Become Progressively Inactivated And Require Two Different Rad51-Mediated Pathways For Restart And Repair. *Mol Cell,* 37**,** 492-502.

Petrakis, T. G., Komseli, E. S., Papaioannou, M., Vougas, K., Polyzos, A., Myrianthopoulos, V., Mikros, E., Trougakos, I. P., Thanos, D., Branzei, D., Townsend, P. & Gorgoulis, V. G. 2016. Exploring And Exploiting The Systemic Effects Of Deregulated Replication Licensing. *Semin Cancer Biol,* 37-38**,** 3-15.

Petryk, N., Kahli, M., D'aubenton-Carafa, Y., Jaszczyszyn, Y., Shen, Y., Silvain, M., Thermes, C., Chen, C. L. & Hyrien, O. 2016. Replication Landscape Of The Human Genome. *Nat Commun,* 7**,** 10208.

Picard, F., Cadoret, J. C., Audit, B., Arneodo, A., Alberti, A., Battail, C., Duret, L. & Prioleau, M. N. 2014. The Spatiotemporal Program Of DNA Replication Is Associated With Specific Combinations Of Chromatin Marks In Human Cells. *Plos Genet,* 10**,** E1004282.

Pomerantz, R. T. & O'donnell, M. 2007. Replisome Mechanics: Insights Into A Twin DNA Polymerase Machine. *Trends Microbiol,* 15**,** 156-64.

Prioleau, M. N. & Macalpine, D. M. 2016. DNA Replication Origins-Where Do We Begin? *Genes Dev,* 30**,** 1683-97.

Prokhorova, T. A. & Blow, J. J. 2000. Sequential Mcm/P1 Subcomplex Assembly Is Required To Form A Heterohexamer With Replication Licensing Activity. *J Biol Chem,* 275**,** 2491-8.

Pruitt, S. C., Bailey, K. J. & Freeland, A. 2007. Reduced Mcm2 Expression Results In Severe Stem/Progenitor Cell Deficiency And Cancer. *Stem Cells,* 25**,** 3121-32.

Raghunandan, M., Chaudhury, I., Kelich, S. L., Hanenberg, H. & Sobeck, A. 2015. Fancd2, Fancj And Brca2 Cooperate To Promote Replication Fork Recovery Independently Of The Fanconi Anemia Core Complex. *Cell Cycle,* 14**,** 342-53.

Rappold, I., Iwabuchi, K., Date, T. & Chen, J. 2001. Tumor Suppressor P53 Binding Protein 1 (53bp1) Is Involved In DNA Damage-Signaling Pathways. *J Cell Biol,* 153**,** 613-20.

Ray Chaudhuri, A., Hashimoto, Y., Herrador, R., Neelsen, K. J., Fachinetti, D., Bermejo, R., Cocito, A., Costanzo, V. & Lopes, M. 2012. Topoisomerase I Poisoning Results In Parp-Mediated Replication Fork Reversal. *Nat Struct Mol Biol,* 19**,** 417-23.

Remus, D., Beuron, F., Tolun, G., Griffith, J. D., Morris, E. P. & Diffley, J. F. 2009. Concerted Loading Of Mcm2-7 Double Hexamers Around DNA During DNA Replication Origin Licensing. *Cell,* 139**,** 719-30.

Rhind, N. & Gilbert, D. M. 2013. DNA Replication Timing. *Cold Spring Harb Perspect Biol,* 5**,** A010132.

Ritzi, M., Baack, M., Musahl, C., Romanowski, P., Laskey, R. A. & Knippers, R. 1998. Human Minichromosome Maintenance Proteins And Human Origin Recognition Complex 2 Protein On Chromatin. *J Biol Chem,* 273**,** 24543-9.

Rogakou, E. P., Boon, C., Redon, C. & Bonner, W. M. 1999. Megabase Chromatin Domains Involved In DNA Double-Strand Breaks In Vivo. *J Cell Biol,* 146**,** 905-16.

Rosin, M. P. & German, J. 1985. Evidence For Chromosome Instability In Vivo In Bloom Syndrome: Increased Numbers Of Micronuclei In Exfoliated Cells. *Hum Genet,* 71**,** 187-91.

Rowles, A., Tada, S. & Blow, J. J. 1999. Changes In Association Of The Xenopus Origin Recognition Complex With Chromatin On Licensing Of Replication Origins. *J Cell Sci,* 112 (Pt 12)**,** 2011-8.

Ryba, T., Hiratani, I., Lu, J., Itoh, M., Kulik, M., Zhang, J., Schulz, T. C., Robins, A. J., Dalton, S. & Gilbert, D. M. 2010. Evolutionarily Conserved Replication Timing Profiles Predict Long-Range Chromatin Interactions And Distinguish Closely Related Cell Types. *Genome Res,* 20**,** 761-70.

Saini, N., Ramakrishnan, S., Elango, R., Ayyar, S., Zhang, Y., Deem, A., Ira, G., Haber, J. E., Lobachev, K. S. & Malkova, A. 2013. Migrating Bubble During Break-Induced Replication Drives Conservative DNA Synthesis. *Nature,* 502**,** 389-92.

Sale, J. E., Lehmann, A. R. & Woodgate, R. 2012. Y-Family DNA Polymerases And Their Role In Tolerance Of Cellular DNA Damage. *Nat Rev Mol Cell Biol,* 13**,** 141-52.

Samel, S. A., Fernandez-Cid, A., Sun, J., Riera, A., Tognetti, S., Herrera, M. C., Li, H. & Speck, C. 2014. A Unique DNA Entry Gate Serves For Regulated Loading Of The Eukaryotic Replicative Helicase Mcm2-7 Onto DNA. *Genes Dev,* 28**,** 1653-66.

Sancar, A., Lindsey-Boltz, L. A., Unsal-Kacmaz, K. & Linn, S. 2004. Molecular Mechanisms Of Mammalian DNA Repair And The DNA Damage Checkpoints. *Annu Rev Biochem,* 73**,** 39-85.

Sanchez, I. & Dynlacht, B. D. 2005. New Insights Into Cyclins, Cdks, And Cell Cycle Control. *Semin Cell Dev Biol,* 16**,** 311-21.

Santamaria, D., Barriere, C., Cerqueira, A., Hunt, S., Tardy, C., Newton, K., Caceres, J. F., Dubus, P., Malumbres, M. & Barbacid, M. 2007. Cdk1 Is Sufficient To Drive The Mammalian Cell Cycle. *Nature,* 448**,** 811-5.

Santocanale, C. & Diffley, J. F. 1996. Orc- And Cdc6-Dependent Complexes At Active And Inactive Chromosomal Replication Origins In Saccharomyces Cerevisiae. *Embo J,* 15**,** 6671-9.

Sartori, A. A., Lukas, C., Coates, J., Mistrik, M., Fu, S., Bartek, J., Baer, R., Lukas, J. & Jackson, S. P. 2007. Human Ctip Promotes DNA End Resection. *Nature,* 450**,** 509-14.

Schaarschmidt, D., Baltin, J., Stehle, I. M., Lipps, H. J. & Knippers, R. 2004. An Episomal Mammalian Replicon: Sequence-Independent Binding Of The Origin Recognition Complex. *Embo J,* 23**,** 191-201.

Schlacher, K., Wu, H. & Jasin, M. 2012. A Distinct Replication Fork Protection Pathway Connects Fanconi Anemia Tumor Suppressors To Rad51-Brca1/2. *Cancer Cell,* 22**,** 106-16.

Schultz, L. B., Chehab, N. H., Malikzay, A. & Halazonetis, T. D. 2000. P53 Binding Protein 1 (53bp1) Is An Early Participant In The Cellular Response To DNA Double-Strand Breaks. *J Cell Biol,* 151**,** 1381-90.

Serrano, M., Hannon, G. J. & Beach, D. 1993. A New Regulatory Motif In Cell-Cycle Control Causing Specific Inhibition Of Cyclin D/Cdk4. *Nature,* 366**,** 704-7.

Sherr, C. J. & Roberts, J. M. 1999. Cdk Inhibitors: Positive And Negative Regulators Of G1-Phase Progression. *Genes Dev,* 13**,** 1501-12.

Sheu, Y. J. & Stillman, B. 2006. Cdc7-Dbf4 Phosphorylates Mcm Proteins Via A Docking Site-Mediated Mechanism To Promote S Phase Progression. *Mol Cell,* 24**,** 101-13.

Shima, N., Alcaraz, A., Liachko, I., Buske, T. R., Andrews, C. A., Munroe, R. J., Hartford, S. A., Tye, B. K. & Schimenti, J. C. 2007. A Viable Allele Of Mcm4 Causes Chromosome Instability And Mammary Adenocarcinomas In Mice. *Nat Genet,* 39**,** 93-8.

Shreeram, S., Sparks, A., Lane, D. P. & Blow, J. J. 2002. Cell Type-Specific Responses Of Human Cells To Inhibition Of Replication Licensing. *Oncogene,* 21**,** 6624-32.

Siow, C. C., Nieduszynska, S. R., Muller, C. A. & Nieduszynski, C. A. 2012. Oridb, The DNA Replication Origin Database Updated And Extended. *Nucleic Acids Res,* 40**,** D682-6.

Smith, D. J. & Whitehouse, I. 2012. Intrinsic Coupling Of Lagging-Strand Synthesis To Chromatin Assembly. *Nature,* 483**,** 434-8.

Smogorzewska, A., Matsuoka, S., Vinciguerra, P., Mcdonald, E. R., 3rd, Hurov, K. E., Luo, J., Ballif, B. A., Gygi, S. P., Hofmann, K., D'andrea, A. D. & Elledge, S. J. 2007. Identification Of The Fanci Protein, A Monoubiquitinated Fancd2 Paralog Required For DNA Repair. *Cell,* 129**,** 289-301.

Speck, C., Chen, Z., Li, H. & Stillman, B. 2005. Atpase-Dependent Cooperative Binding Of Orc And Cdc6 To Origin DNA. *Nat Struct Mol Biol,* 12**,** 965-71.

Spruck, C. H., Won, K. A. & Reed, S. I. 1999. Deregulated Cyclin E Induces Chromosome Instability. *Nature,* 401**,** 297-300.

Stavropoulos, D. J., Bradshaw, P. S., Li, X., Pasic, I., Truong, K., Ikura, M., Ungrin, M. & Meyn, M. S. 2002. The Bloom Syndrome Helicase Blm Interacts With Trf2 In Alt Cells And Promotes Telomeric DNA Synthesis. *Hum Mol Genet,* 11**,** 3135-44.

Sun, H., Karow, J. K., Hickson, I. D. & Maizels, N. 1998. The Bloom's Syndrome Helicase Unwinds G4 DNA. *J Biol Chem,* 273**,** 27587-92.

Sun, J., Evrin, C., Samel, S. A., Fernandez-Cid, A., Riera, A., Kawakami, H., Stillman, B., Speck, C. & Li, H. 2013. Cryo-Em Structure Of A Helicase Loading Intermediate Containing Orc-Cdc6-Cdt1-Mcm2-7 Bound To DNA. *Nat Struct Mol Biol,* 20**,** 944-51.

Takahashi, T. S., Wigley, D. B. & Walter, J. C. 2005. Pumps, Paradoxes And Ploughshares: Mechanism Of The Mcm2-7 DNA Helicase. *Trends Biochem Sci,* 30**,** 437-44.

Takai, H., Smogorzewska, A. & De Lange, T. 2003. DNA Damage Foci At Dysfunctional Telomeres. *Curr Biol,* 13**,** 1549-56.

Takeda, D. Y., Parvin, J. D. & Dutta, A. 2005. Degradation Of Cdt1 During S Phase Is Skp2-Independent And Is Required For Efficient Progression Of Mammalian Cells Through S Phase. *J Biol Chem,* 280**,** 23416-23.

Tanaka, S. & Diffley, J. F. 2002. Interdependent Nuclear Accumulation Of Budding Yeast Cdt1 And Mcm2-7 During G1 Phase. *Nat Cell Biol,* 4**,** 198-207.

Tanaka, S., Nakato, R., Katou, Y., Shirahige, K. & Araki, H. 2011. Origin Association Of Sld3, Sld7, And Cdc45 Proteins Is A Key Step For Determination Of Origin-Firing Timing. *Curr Biol,* 21**,** 2055-63.

Tanaka, S., Umemori, T., Hirai, K., Muramatsu, S., Kamimura, Y. & Araki, H. 2007. Cdk-Dependent Phosphorylation Of Sld2 And Sld3 Initiates DNA Replication In Budding Yeast. *Nature,* 445**,** 328-32.

Teer, J. K., Machida, Y. J., Labit, H., Novac, O., Hyrien, O., Marheineke, K., Zannis-Hadjopoulos, M. & Dutta, A. 2006. Proliferating Human Cells Hypomorphic For Origin Recognition Complex 2 And Pre-Replicative Complex Formation Have A Defect In P53 Activation And Cdk2 Kinase Activation. *J Biol Chem,* 281**,** 6253-60.

Teng, S. C. & Zakian, V. A. 1999. Telomere-Telomere Recombination Is An Efficient Bypass Pathway For Telomere Maintenance In Saccharomyces Cerevisiae. *Mol Cell Biol,* 19**,** 8083-93.

Theis, J. F., Irene, C., Dershowitz, A., Brost, R. L., Tobin, M. L., Di Sanzo, F. M., Wang, J. Y., Boone, C. & Newlon, C. S. 2010. The DNA Damage Response Pathway Contributes To The Stability Of Chromosome Iii Derivatives Lacking Efficient Replicators. *Plos Genet,* 6**,** E1001227.

Toledo, L. I., Altmeyer, M., Rask, M. B., Lukas, C., Larsen, D. H., Povlsen, L. K., Bekker-Jensen, S., Mailand, N., Bartek, J. & Lukas, J. 2013. ATR Prohibits Replication Catastrophe By Preventing Global Exhaustion Of Rpa. *Cell,* 155**,** 1088-103.

Tumini, E., Barroso, S., Calero, C. P. & Aguilera, A. 2016. Roles Of Human Pold1 And Pold3 In Genome Stability. *Sci Rep,* 6**,** 38873.

Tye, B. K. 1999. Mcm Proteins In DNA Replication. *Annu Rev Biochem,* 68**,** 649-86.

Varma, D., Chandrasekaran, S., Sundin, L. J., Reidy, K. T., Wan, X., Chasse, D. A., Nevis, K. R., Deluca, J. G., Salmon, E. D. & Cook, J. G. 2012. Recruitment Of The Human Cdt1 Replication Licensing Protein By The Loop Domain Of Hec1 Is Required For Stable Kinetochore-Microtubule Attachment. *Nat Cell Biol,* 14**,** 593-603.

Vashee, S., Cvetic, C., Lu, W., Simancek, P., Kelly, T. J. & Walter, J. C. 2003. Sequence-Independent DNA Binding And Replication Initiation By The Human Origin Recognition Complex. *Genes Dev,* 17**,** 1894-908.

Vashee, S., Simancek, P., Challberg, M. D. & Kelly, T. J. 2001. Assembly Of The Human Origin Recognition Complex. *J Biol Chem,* 276**,** 26666-73.

Vassilev, L. T. 2006. Cell Cycle Synchronization At The G2/M Phase Border By Reversible Inhibition Of Cdk1. *Cell Cycle,* 5**,** 2555-6.

Vassilev, L. T., Tovar, C., Chen, S., Knezevic, D., Zhao, X., Sun, H., Heimbrook, D. C. & Chen, L. 2006. Selective Small-Molecule Inhibitor Reveals Critical Mitotic Functions Of Human Cdk1. *Proc Natl Acad Sci U S A,* 103**,** 10660-5.

Vaziri, C., Saxena, S., Jeon, Y., Lee, C., Murata, K., Machida, Y., Wagle, N., Hwang, D. S. & Dutta, A. 2003. A P53-Dependent Checkpoint Pathway Prevents Rereplication. *Mol Cell,* 11**,** 997-1008.

Villa, F., Simon, A. C., Ortiz Bazan, M. A., Kilkenny, M. L., Wirthensohn, D., Wightman, M., Matak-Vinkovic, D., Pellegrini, L. & Labib, K. 2016. Ctf4 Is A Hub In The Eukaryotic Replisome That Links Multiple Cip-Box Proteins To The Cmg Helicase. *Mol Cell,* 63**,** 385-96.

Wang, L. H., Mayer, B., Stemmann, O. & Nigg, E. A. 2010. Centromere DNA Decatenation Depends On Cohesin Removal And Is Required For Mammalian Cell Division. *J Cell Sci,* 123**,** 806-13.

Wang, W. 2007. Emergence Of A DNA-Damage Response Network Consisting Of Fanconi Anaemia And Brca Proteins. *Nat Rev Genet,* 8**,** 735-48.

Ward, I. M., Minn, K., Jorda, K. G. & Chen, J. 2003a. Accumulation Of Checkpoint Protein 53bp1 At DNA Breaks Involves Its Binding To Phosphorylated Histone H2ax. *J Biol Chem,* 278**,** 19579-82.

Ward, I. M., Minn, K., Van Deursen, J. & Chen, J. 2003b. P53 Binding Protein 53bp1 Is Required For DNA Damage Responses And Tumor Suppression In Mice. *Mol Cell Biol,* 23**,** 2556-63.

Webb, C. J., Wu, Y. & Zakian, V. A. 2013. DNA Repair At Telomeres: Keeping The Ends Intact. *Cold Spring Harb Perspect Biol,* 5.

Weber, J. D., Taylor, L. J., Roussel, M. F., Sherr, C. J. & Bar-Sagi, D. 1999. Nucleolar Arf Sequesters Mdm2 And Activates P53. *Nat Cell Biol,* 1**,** 20-6.

Whitby, M. C., Osman, F. & Dixon, J. 2003. Cleavage Of Model Replication Forks By Fission Yeast Mus81-Eme1 And Budding Yeast Mus81-Mms4. *J Biol Chem,* 278**,** 6928-35.

Widrow, R. J., Hansen, R. S., Kawame, H., Gartler, S. M. & Laird, C. D. 1998. Very Late DNA Replication In The Human Cell Cycle. *Proc Natl Acad Sci U S A,* 95**,** 11246-50.

Wilmes, G. M., Archambault, V., Austin, R. J., Jacobson, M. D., Bell, S. P. & Cross, F. R. 2004. Interaction Of The S-Phase Cyclin Clb5 With An "Rxl" Docking Sequence In The Initiator Protein Orc6 Provides An Origin-Localized Replication Control Switch. *Genes Dev,* 18**,** 981-91.

Woodward, A. M., Gohler, T., Luciani, M. G., Oehlmann, M., Ge, X., Gartner, A., Jackson, D. A. & Blow, J. J. 2006. Excess Mcm2-7 License Dormant Origins Of Replication That Can Be Used Under Conditions Of Replicative Stress. *The Journal Of Cell Biology,* 173**,** 673-83.

Xia, Z., Morales, J. C., Dunphy, W. G. & Carpenter, P. B. 2001. Negative Cell Cycle Regulation And DNA Damage-Inducible Phosphorylation Of The Brct Protein 53bp1. *J Biol Chem,* 276**,** 2708-18.

Xouri, G., Squire, A., Dimaki, M., Geverts, B., Verveer, P. J., Taraviras, S., Nishitani, H., Houtsmuller, A. B., Bastiaens, P. I. & Lygerou, Z. 2007. Cdt1 Associates Dynamically With Chromatin Throughout G1 And Recruits Geminin Onto Chromatin. *Embo J,* 26**,** 1303-14.

Yamamoto, K. N., Kobayashi, S., Tsuda, M., Kurumizaka, H., Takata, M., Kono, K., Jiricny, J., Takeda, S. & Hirota, K. 2011. Involvement Of Slx4 In Interstrand Cross-Link Repair Is Regulated By The Fanconi Anemia Pathway. *Proc Natl Acad Sci U S A,* 108**,** 6492-6.

Yan, S. & Michael, W. M. 2009. Topbp1 And DNA Polymerase-Alpha Directly Recruit The 9-1-1 Complex To Stalled DNA Replication Forks. *J Cell Biol,* 184**,** 793-804.

Yardimci, H. & Walter, J. C. 2014. Prereplication-Complex Formation: A Molecular Double Take? *Nat Struct Mol Biol,* 21**,** 20-5.

Yeeles, J. T., Deegan, T. D., Janska, A., Early, A. & Diffley, J. F. 2015. Regulated Eukaryotic DNA Replication Origin Firing With Purified Proteins. *Nature,* 519**,** 431-5.

Yeeles, J. T., Janska, A., Early, A. & Diffley, J. F. 2017. How The Eukaryotic Replisome Achieves Rapid And Efficient DNA Replication. *Mol Cell,* 65**,** 105-116.

Yeeles, J. T., Poli, J., Marians, K. J. & Pasero, P. 2013. Rescuing Stalled Or Damaged Replication Forks. *Cold Spring Harb Perspect Biol,* 5**,** A012815.

Yekezare, M., Gomez-Gonzalez, B. & Diffley, J. F. 2013. Controlling DNA Replication Origins In Response To DNA Damage - Inhibit Globally, Activate Locally. *J Cell Sci,* 126**,** 1297-306.

Ying, S., Minocherhomji, S., Chan, K. L., Palmai-Pallag, T., Chu, W. K., Wass, T., Mankouri, H. W., Liu, Y. & Hickson, I. D. 2013. Mus81 Promotes Common Fragile Site Expression. *Nat Cell Biol,* 15**,** 1001-7.

Zegerman, P. & Diffley, J. F. 2007. Phosphorylation Of Sld2 And Sld3 By Cyclin-Dependent Kinases Promotes DNA Replication In Budding Yeast. *Nature,* 445**,** 281-5.

Zellweger, R., Dalcher, D., Mutreja, K., Berti, M., Schmid, J. A., Herrador, R., Vindigni, A. & Lopes, M. 2015. Rad51-Mediated Replication Fork Reversal Is A Global Response To Genotoxic Treatments In Human Cells. *J Cell Biol,* 208**,** 563-79.

Zhang, Y. & Xiong, Y. 2001. Control Of P53 Ubiquitination And Nuclear Export By Mdm2 And Arf. *Cell Growth Differ,* 12**,** 175-86.

Zhu, W., Chen, Y. & Dutta, A. 2004. Rereplication By Depletion Of Geminin Is Seen Regardless Of P53 Status And Activates A G2/M Checkpoint. *Mol Cell Biol,* 24**,** 7140-50.

Zhu, W. & Dutta, A. 2006. An ATR- And Brca1-Mediated Fanconi Anemia Pathway Is Required For Activating The G2/M Checkpoint And DNA Damage Repair Upon Rereplication. *Mol Cell Biol,* 26**,** 4601-11.

Zimmermann, M., Lottersberger, F., Buonomo, S. B., Sfeir, A. & De Lange, T. 2013. 53bp1 Regulates Dsb Repair Using Rif1 To Control 5' End Resection. *Science,* 339**,** 700-4.